The Dissertation Committee for Matthew Lyall Evans Certifies that this is the approved version of the following dissertation:

The Authority of Pleasure and Pain: Moral Psychology in Plato's *Philebus*

Committee:

Stephen White, Supervisor

David Sosa, Co-Supervisor

Jim Hankinson

Paul Woodruff

Brian Leiter

Terence Irwin

The Authority of Pleasure and Pain: Moral Psychology in Plato's *Philebus*

by

Matthew Lyall Evans, B.A., M.A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2004

Dedication

for my parents

The Authority of Pleasure and Pain: Moral Psychology in Plato's *Philebus*

Publication No._____

Matthew Lyall Evans, Ph.D.

The University of Texas at Austin, 2004

Supervisors: Stephen White and Ernest D. Sosa

I interpret, analyze, and defend Plato's views about the nature and value of pleasure and pain, with special attention to the way he develops these views in the *Philebus*. The core of Plato's position, I argue, is that pleasure as such is not a bearer of final value, but a fallible mode of perceiving final value. If Plato is right, then the claim that pleasure as such is the ethical goal is akin to the claim that belief as such is the epistemological goal. What makes this version of anti-hedonism exciting, to my mind, is that it can concede that pleasure has authority in itself—in that it always gives us a reason to act as it bids—without conceding that pleasure is *good* in itself. Moreover, an account of this sort is refreshingly free of the asceticism that pervades so much of both ancient and modern anti-hedonism, including some of Plato's own earlier work. After establishing that Plato advances this view in the *Philebus*, I argue that the view itself yields genuine insight into the psychological role and ethical status of pleasure and pain.

Table of Contents

0. Introduction

The *Philebus* is an unlovely creature. It has little of the fine style and rigorous order we find in Plato's much-celebrated middle-period works. In Bury's 1897 edition of the dialogue, he aptly likened it "to a gnarled and knotted old oak-tree, abounding in unexpected humps and shoots, which sadly mar its symmetry as compared with the fair cypress-trees and stately pines by whose side it stands in the grove of Academe" (ix). Yet despite what Bury calls its "harsh and rugged style" and its "jagged and distorted composition," the *Philebus* is one of Plato's richest works. It is packed from top to bottom with forceful, beguiling arguments that both demand and reward close attention. In the time I have spent negotiating its peculiar twists and turns, I have steadily become convinced that the dialogue's forays into ethical theory, philosophy of mind, and moral psychology are more sophisticated and penetrating than anyone has yet fully appreciated. The overall goal of my dissertation is to bring Plato's neglected achievement to light, and to encourage historically-minded ethical philosophers to take a detour from the prettier, better-kept path and risk a hike through the *Philebus*.

My specific goal is to show that Plato develops a surprisingly radical anti-hedonist position in the dialogue, and that the argument supporting it is coherent, forceful, and philosophically appealing in its own right. His argument is best understood, I think, as a dialectical response to a view held by Eudoxus of Cnidus, a famous

astronomer and mathematician who was also a close colleague of Plato's at the Academy.[1] According to Aristotle, Eudoxus advanced an argument to the effect that pleasure is the only thing worth choosing for its own sake. His argument runs as follows in Aristotle's text:

> What is most choiceworthy is what we choose neither because of, nor for the sake of, anything else. And it is agreed that this is pleasure, since we never ask anyone what his end is in being pleased, on the assumption that pleasure is choiceworthy in itself. (*Nicomachean Ethics* X, 1172b19-23)

Though Aristotle's report is compressed, the thrust of Eudoxus' argument is fairly clear. Eudoxus wants to claim, first, that our practice of giving reasons for our actions shows that pleasure is something we choose only for its own sake; and second, that pleasure, since it is something we choose only for its own sake, must be the most choiceworthy thing there is.

The argument has some obvious weaknesses. It is not clear, first of all, that pleasure is something we choose only for its own sake. Sometimes we seek enjoyment to refresh ourselves before tackling a difficult project, or to expand our understanding of an aesthetic field. Our reasons for seeking pleasure need not give out at the point of getting it. Moreover, even if pleasure were something we pursue only for its own sake, it does not follow from this that pleasure is the *only* thing we pursue only for its own sake. Eudoxus has not adequately ruled out the possibility that there are other things that have this status, such as intellectual enrichment or the raising of children. To establish the

[1] On Eudoxus' hedonism, see Gosling and Taylor (1982, 157-164, 255-283).

stronger point, he would have to argue in addition that these other activities are in fact undertaken for the sake of pleasure. And there is no good reason to think that any such argument, even if it were in the offing, would be plausible. In short, Eudoxus has shown neither that pleasure is something we choose only for its own sake, nor that pleasure as such is the *only* thing we choose only for its own sake.

Still, there is something about the argument that seems right. If it is restricted to the claim that our standards of practical reason-giving imply that pleasure is something we choose for its own sake—though *not always only* for its own sake—then his argument can be seen to carry some weight. When you justify your action by claiming that you just enjoy doing that sort of thing, your answer does seem to be *prima facie* sufficient. To ask what further end you have in enjoying yourself seems unreasonable if not bizarre. As Elizabeth Anscombe (1957, 77) writes, "'it's pleasant' is an adequate answer to 'What's the good of it?' or 'What do you want that for?' I.e., the chain of 'Why's' comes to an end with this answer." Thomas Nagel (1986, 156) makes a similar claim: "almost everyone takes the avoidance of his own pain and the promotion of his own pleasure as subjective reasons for action in a fairly simple way: they are not backed up by any further reasons."[2] What Eudoxus seems to have grasped, then, is that a thing's being pleasant to me gives me, in itself, a reason to do or to have that thing. And from this he might infer that pleasure is, in every case, something to be pursued for its own sake.

[2] Nagel (1986, 156-162) goes on to argue that the authority of these subjective reasons is so entrenched that "no objective view we can attain could possibly overrule" it.

If my reading of the *Philebus* is correct, then Plato effectively dedicates the dialogue to rejecting this last inference. In his view, no pleasure is ever worth choosing for its own sake, and we commit a serious error whenever we pursue any pleasure as such. This is the sense in which Plato's anti-hedonism is radical, rather than merely reformist. He does not limit himself to the claim that *not only* or *not all* pleasures are final goods; he insists that *no* pleasures are final goods, and that one should *never* pursue them for their own sake. As I understand it, Plato's diagnosis of the Eudoxan argument is that it trades on an ambiguity between the *attitude* of pleasure and the *content of* this attitude—between *the event of enjoyment* and *what is enjoyed*. Even if it is granted that our taking pleasure in some activity gives us a reason to pursue that activity, it does not follow from this that the pleasure-taking itself, rather than the activity in which it is taken, is what we have a reason to pursue. Once we sniff out this mistake, Plato thinks, we will be in a position to see that pleasure's own reason-giving authority indicates that it itself is *never* something we have a reason to pursue.

In support of this move, Plato develops and defends two central claims: first, that pleasures are not mere sensations or "raw feels," but are attitudes rich with intentional content; and second, that pleasure-attitudes have the representational function of picking out properties, events, or states of affairs that we should appreciate as good. On this sort of view, pleasures are quasi-perceptual states that are uniquely and intrinsically sensitive to value, since their constitutive aim is to represent objects worthy of ultimate pursuit. Hence they have a kind of *guiding* role to play in our deliberative economy, telling us, in

effect, what to shoot for in our actions.[3] Pleasure-taking, then, is a way to select practical targets, where these targets are selected not *qua* pleasant, but *qua good*. If this is right, then the Eudoxan argument—even in its weakened form—is badly misguided. For it hinges on the subtle but clear mistake of confusing guides and goals. The ethical error of aiming for pleasure as such is thus akin to the epistemological error of aiming at *belief* as such: neither target is suitable as an ultimate end because the value of each crucially depends on the value of its content.[4] Just as we do not want beliefs with false content, we do not want *pleasures* with *bad* content. Eudoxus, and those who follow him, do not fully appreciate this. They make the mistake of supposing that, if we always have a reason to do as pleasure bids, then we always have a reason to aim for pleasure itself. But if Plato is right, pleasure's own intrinsic reason-giving function makes it unsuitable to serve as such an end.[5]

My argument proceeds in three stages. In chapter 1 I establish that Plato advocates a broadly intentionalist theory of pleasure and pain. My burden here is to show not only that Plato is interested in developing a substantive theory of this sort, but

---

[3] For a contemporary account of pleasure and pain that is remarkably close to the one I am attributing to Plato here, see Millgram (1993, 397-400). One difference between the two accounts, worth noting here, is that Millgram thinks of pleasure as a *feeling* rather than an *attitude*. Unlike Plato, he is especially sensitive to pleasure's status as a *qualitative* psychological state.

[4] I owe this way of putting the point to Millgram (1993, 400-401): "Practical reasoning tends to take one from a position of lesser pleasure to a position of greater pleasure. … Some philosophers have noticed this tendency, and concluded that pleasure is one's sole and necessary goal. In this they could not be more mistaken. Hedonists err in roughly the way that someone who thinks that the goal of enquiry is to maximize conviction might err. Normally, one's enquiries tend to take one from a position of lesser conviction to a position of greater conviction… However, … one's goal is not conviction: one's goal is truth. Conviction is epistemically important as a guide to truth, but conviction *per se* is not the object of my efforts."

[5] It is important to notice that Plato's anti-hedonism has no truck with *ascetic* anti-hedonism, the view that pleasure is always to be avoided. If I am right, Plato is perfectly comfortable with the idea that, other things being equal, we should do as pleasure bids. What he rejects is the idea that we should do as pleasure bids *because what pleasure bids is pleasant*.

also that the theory he develops supports his anti-hedonism in the way I have suggested. To this end I argue that Plato's various remarks can be woven into a coherent account, according to which all pleasures are content-bearing attitudes that consist in a restoration of psychological harmony or equilibrium. More intuitively, his view is that to be pleased is to *find one's mental balance* with respect to one's concerns. This general intentionalist position naturally leads him to the view that pleasures can be false, since if pleasures can have intentional content, then they can also have false intentional content. I focus on this view in chapter 2, arguing that Plato succeeds in showing that at least some pleasures can be false—provided that what it is to be false is to have false content. But Plato wants to show more than this, I argue. In his view, pleasures are like beliefs in that bad pleasures are bad *just because* they are false. After arguing that Plato cannot support this latter claim without an improved account of what it is for a pleasure to be false, I suggest that the ingredients for such an account can be found in a later and often-overlooked part of the *Philebus*. Here Plato gives every indication of thinking that what makes a pleasure false is not its having false content, but rather its having content that fails to represent an object of genuine value. And this, I argue, is precisely the account he needs to vindicate his idea that bad pleasures are bad because they are false.

In the third and final chapter, I argue that Plato supports his explicitly adopted anti-hedonism with the very claims I attribute to him in chapters 1 and 2. As I understand it, his goal is to show that pleasure, understood now as a kind of attitude, cannot be a bearer of final value because it is the sort of thing that intrinsically aims at something

other than itself. Belief-attitudes aim at the truth in general, and pleasure-attitudes aim at the truth *about what is good*. For this reason, Plato holds that neither beliefs nor pleasures are suitable objects of ultimate pursuit. A number of commentators have challenged this interpretation, however, and I devote a large part of the chapter to addressing their views. Their most pressing objection appeals to an argument, advanced early in the dialogue, which seems to conclude that at least some pleasures are choiceworthy in themselves. I defend an alternative interpretation of this argument, according to which it establishes that the good human life will be a pleasant one, but not that the pleasantness in this life is even part of what *makes* it good. Though Plato is an anti-hedonist, in my view, he is not an ascetic. He consistently and forcefully denies that the pleasureless life is the best one for us. But he also denies that the pleasant life is best for us *because* it is pleasant. Whatever value pleasure has, he thinks, derives entirely from the value of its objects.

Before plunging into my argument, let me take a moment to warn the reader that I confine myself rather narrowly to the task of interpreting the *Philebus*. With a few minor exceptions, I do not draw on Plato's other works in my attempt to make sense of the arguments in this one. Nor do I bring his view into dialogue with other ancient authors, such as Aristotle or Epicurus, who deal with similar topics. Though I believe such an enterprise would be highly illuminating, I also believe that a more tightly focused approach has its own virtues. For one thing, it forces the interpreter to grapple with the arguments as they stand in the text, with no resources other than what the text itself

provides. By limiting the scope and ambition of my work in this way, I hope it can serve as a sound springboard for more wide-ranging and comprehensive work in the future.

1. A Theory of Pleasure and Pain

1.1. Plato's Theory: An Introduction

The *Philebus* contains Plato's most extensive reflections on the nature of pleasure and pain, but it is not obvious that these reflections express, even in outline, a *substantive theory* of pleasure and pain. To qualify as substantive, a theory of X would have to pick out some non-disjunctive X-making property that is possessed by all and only Xs. Yet as several commentators have noticed, Plato's remarks in the *Philebus* seem to fall well short of satisfying this *unity constraint* on substantive theories.[6] Though he insists that pleasure and pain come in a dizzying array of different types, and though he discusses several of them at length, he never attempts to show that each of these different types bears some common property. Nor does he provide a fully fleshed-out account of what that pleasure-making or pain-making property might be.

From this it is tempting to conclude, as Gosling and Taylor (1982, 135-41) do, that Plato is simply not interested in building a substantive theory. On their reading—which I will call the *no-theory view*—Plato does not even try to meet the unity constraint because he holds that at least some forms of pleasure have nothing substantial

[6] Gosling and Taylor (1982, 140) and Tuozzo (1996, 495-498) both notice and take careful account of this apparent deficiency.

in common.[7] What motivates the no-theory view is the thought that, in Plato's view, it is pleasure's radical complexity that makes hedonism an incoherent doctrine. If there is no substantive unity to pleasure, the thinking goes, then those who claim that "pleasure is the good" are making a mistake. For they presuppose that there is some substantive property, shared by all and only pleasures, the possession of which makes each pleasure good. The no-theory advocates hold that this is an important step in Plato's critique of hedonism, and that any attempt to find in the *Philebus* a substantive theory of pleasure is, for this reason, misguided. In their view, such an attempt is worse than a waste of effort; it reflects a failure to appreciate Plato's overall strategy.

My goal in this chapter is to show that the no-theory view of Plato's project is wrong, and that his remarks on the nature of pleasure and pain can together yield a theory that satisfies the unity constraint.[8] I begin by establishing that Plato thinks of pleasure and pain as genuine mental kinds, and that Plato's view of kinds in general commits him to the idea that all the members of a specific kind must have some non-disjunctive property in common. From this I infer that Plato is committed, at least in principle, to making his account of pleasure and pain satisfy the unity constraint. But in order to accomplish this, Plato needs to show that there is some pleasure-making property that belongs exclusively and exhaustively to the many different types of pleasure he discusses. Since the no-theory view derives most of its force from Plato's apparent failure in this task, my burden is to establish that Plato's failure is merely apparent. To that end, I

[7] See also Gosling (1975, 141-142). Hampton (1990, 73) joins Gosling and Taylor in endorsing this view and Carone (2000, 267n19) is sympathetic.

reconstruct what I take to be the only account of pleasure and pain on offer in the *Philebus*—an account that I call the *equilibrium model* (EM). According to EM, roughly, every pleasure is a process by which the internal harmony of an animal is established (or re-established) and every pain is a process by which an animal's internal harmony is dissolved or destroyed. I argue that a close, critical examination of Plato's account of bodily pleasure and pain shows how EM can be applied successfully to each of the other types of pleasure and pain he discusses. Finally—and again, contrary to the no-theory view—I suggest that a fleshed-out version of EM plays a crucial supporting role in Plato's attack on hedonism.

1.2. Unity and Diversity

Any substantive theory of pleasure must pick out some non-disjunctive, non-trivial, and pleasure-making property that belongs to all and only pleasures. Though this requirement might look harmless, particularly to the average Platonist, Plato himself occasionally seems wary of it. In the *Philebus* we find him so impressed by the diversity of hedonic experience that he appears at times to doubt whether any unified theory could possibly do it justice. And this doubt, whether or not Plato seriously entertains it, is a reasonable one. It seems evident that pleasures differ from one another, and that they do so at least partly in virtue of differences in their objects. The object of a pleasure, as I use

[8] Thus I am in broad agreement with Frede (1992, 437-442) and Tuozzo (1996).

the term, is *that about which* we are pleased, *that in which* we take pleasure, or *that which* we enjoy.

The potential diversity of pleasure's objects is staggering. We find ourselves enjoying very different kinds of things—eating a meal, watching a film, hiking a trail, listening to music, completing proofs, daydreaming. And these pleasures have phenomenological differences that seem to reflect differences in their objects. Some pleasures have objects that are well-expressed within the "pleased that p" idiom, but others do not: to be *pleased that* this sculpture is a Boccioni is not the same as to *take pleasure in* this sculpture *as* a Boccioni.[9] There are, in addition, pleasures of anticipation and recollection, whose objects seem to be temporally bound, and pleasure-moods—such as free-floating elation or bliss—that seem to have no distinctive objects whatsoever. All this goes to show that pleasure can come in an enormous variety of different forms, depending on what its objects are. So one might be tempted to think that pleasures ultimately have nothing more in common with each other than their potential objects do.

Plato seems to flirt with this conclusion in the *Philebus*. He shows an acute awareness that pleasures take objects (37a1-10); that pleasures differ from one another in virtue of differences in their objects (12c8-d6); and that some pleasures take objects that can be expressed as propositions (32b9-c6, 40a3-b5). Moreover, at a crucial early stage of the dialogue, Plato seems to endorse the view that pleasures are radically diverse:

> Socrates: As for pleasure, I know [οἶδα] that it is complex [ποικίλον]. And we should, starting from this, reflect upon and examine what sort of nature it

[9] On the difference, see Williams (1959, 66).

has. For pleasure, if we understand it in an unconditional way [ἁπλῶς], is some one thing; but in fact it takes all sorts of forms that are somehow unlike one another [μορφὰς δὲ δήπου παντοίας εἴληφε καί τινα τρόπον ἀνομοίους ἀλλήλαις]. Look, we say that the intemperate person takes pleasure, but that the temperate person takes pleasure in being temperate itself [ἥδεσθαι δὲ καὶ τὸν σωφρονοῦντα αὐτῷ τῷ σωφρονεῖν]. We also say that the mindless person who is full of mindless judgments and hopes takes pleasure, but that the intelligent person takes pleasure in being intelligent itself. How could anyone say of each pair of these pleasures that they are like one another [τούτων τῶν ἡδονῶν ἑκατέρας πῶς ἄν τις ὁμοίας ἀλλήλαις εἶναι λέγων] and not justly be regarded as mindless?

Protarchus: But the reason for this, Socrates, is that these [pleasures] come from opposite things [ἀπ' ἐναντίων], not that they themselves are opposed to one another [οὐ μὴν αὐταί γε ἀλλήλαις ἐναντίαι]. For how could pleasure not be of all things most like [ὁμοιότατον] pleasure, this very thing most like itself? (12c4-e2)

Socrates makes a number of interesting claims in this passage, but only three of them need concern us here: (1) pleasure is "complex" in that it "takes all sorts of forms which are in some way unlike one another" (12c7-8); (2) the differences between the various types of pleasure reflect differences in their objects (12c8-d4); and (3) some types of pleasure are not like one another (12d4-6). (1) and (2) seem fairly straightforward: there are many different types of pleasure, and the differences between these types stem from differences in their objects. But (3) is ambiguous. It can be read to mean either

(3a) that opposed pleasures are not alike *in any way*, or

(3b) that opposed pleasures are not alike *in every way*.

Only (3a) carries a commitment to the claim that some pleasures have nothing in common. (3b) is silent on that issue.

Defenders of the no-theory view opt for (3a), which at first glance seems to be the more natural reading of the text.[10] Socrates does seem to deny that opposed pleasures are alike, and this strongly suggests that he thinks they have nothing in common. (3a) also makes sense of Socrates' subsequent warning against any account that would take Protarchus' cue and "make all the most opposed things one" (πάντα τὰ ἐναντιώτατα ἓν ποιοῦντι, 13a3-4). The "most opposed things" he refers to here are presumably the "pleasures opposed to pleasures" (ἡδονὰς ἡδοναῖς ἐναντίας) that Socrates goes on to mention a bit later (13a4-5). If the no-theory view is right that (3a) is the better interpretation of (3), then Socrates' central claim in this passage is that there can be no unified, non-disjunctive account of all the various types of pleasure, since at least some of these types have nothing in common. On this reading, Protarchus demurs because he thinks that any two pleasures must be alike insofar as they are pleasures (cf. 13c3-5). His objection therefore amounts to a denial of the claim that there are some wholly dissimilar types of pleasure.

But (3a) cannot be right. First of all, it simply does not fit with what Socrates immediately goes on to say in his reply to Protarchus' objection. Instead of defending the view that pleasure is radically diverse, Socrates draws an analogy between the complexity of pleasure and the complexity of color and shape (12e3-13a3). He insists that some colors and some shapes are "most opposed" (ἐναντιώτατον) to one another (12e6,

13a1), yet he explicitly concedes that colors "will not differ insofar as they are colors" (κατά γε αὐτὸ τοῦτο οὐδὲν διοίσει τὸ χρῶμα εἶναι, 12e3-4) and that every shape is "one in genus" (γένει μέν ἐστι πᾶν ἕν, 12e7). It would be exceedingly strange for Socrates to make these concessions and still cling to the view that the "most opposed" colors or shapes have nothing substantive in common.[11] So if the analogy between pleasure and color (or shape) is supposed to extend this far, then (3a) seems to be in trouble as a reading of (3). Socrates expresses no doubt whatsoever that there is a genus of pleasure, just as there is a genus of color and of shape; indeed, he insists that his sole purpose in dividing pleasure into types is to find out "whether the genus as a whole [ὅλον τὸ γένος] should be welcomed" (32c8-d2). So if Plato supposes that every member of a given genus must have at least one common property that explains their place in that genus, then he cannot think that the "opposed pleasures" have nothing substantive in common.

Defenders of the no-theory view might deny the antecedent. They might insist that Plato can classify all pleasures under a single genus without conceding that they all share some non-disjunctive property (other than the trivial property of being a pleasure). But this too is a dead end, at least as a reading of Plato. In the *Philebus* Socrates repeatedly implies that the same name is correctly applied to a number of different things

[10] Cf. Gosling (1975, 77).
[11] In the *Meno* (74b-75a), Socrates uses the same examples to convince Meno that specific differences and oppositions between shapes and colors in no way releases him from having to give a generic account of each. Cf. Irwin (1995, 320-321) and Tuozzo (1996, 496).

only if they share some distinct, non-disjunctive, and non-trivial property.[12] Socrates first articulates this principle in responding to the claim—offered up with much fanfare by Protarchus—that all pleasures are good (13a8). Socrates denies this, urging instead that only some are (13b1). He then asks Protarchus: "what is the same thing inhering in the bad and good pleasures alike [τί οὖν δὴ ταὐτὸν ἐν ταῖς κακαῖς ὁμοίως καὶ ἐν ἀγαθαῖς ἐνὸν] such that you address [προσαγορεύεις] all pleasures as being good" (13b3-5). While this question may initially strike us as unfair, that Socrates asks it at all implies that he holds fast to the relevant principle. In his view, Protarchus is justified in calling a number of things by a single name only if he can specify some feature, shared by all of these things, their possession of which justifies his application of the name. Later, when discussing the topic of desire, Socrates asks a similar version of the same question: "what is the same thing we look at when we call these many different things by a single name" (πρὸς τί ποτε ἄρα ταὐτὸν βλέψαντες οὕτω πολὺ διαφέροντα ταῦθ' ἑνὶ προσαγορεύομεν ὀνόματι, 34e3-4). Once again the implication is that a single name applies correctly to a set of different things only if all of these things share some property that justifies the application of the name. So if Socrates accepts that the various types of pleasure are properly called "pleasures" in virtue of belonging to a single genus, then he must concede that all these types share some property that warrants their all being called "pleasures." If, on the other hand, he holds that "opposed pleasures" have nothing in common, then he must think that he himself makes a serious mistake in calling

[12] This assumption underlies most of the "what is F-ness" questions that Socrates asks of his interlocutors throughout the Socratic dialogues. For a representative example of this, see *Euthyphro* 6e.

them both "pleasures" in the first place. But he shows no sign of acknowledging such a mistake.

Moreover, in his discussion of the so-called "Promethean method" of collection and division, Plato implies that discovering the common property of every member of a genus is just as important an intellectual task as dividing the genus into its proper species (16e10-d2; 17d6-e6). And Plato leaves no doubt that he thinks of pleasure as a genus to which the "Promethean method" can and should be applied (18e3-19a2). Just after Socrates introduces the method, Protarchus suggests that the whole point of introducing it is to bring it to bear on pleasure and intelligence. Speaking to Philebus, he says: "Socrates seems to be asking us whether there are types [εἴδη] of pleasure or not, and how many [ὁπόσα] there are, and of what sort [ὁποῖα] they are" (19b2-3). Socrates agrees. But this is only part of the procedure under discussion. As Socrates understands it, the procedure of collecting and dividing is complete only when one has established both that and how the original item under consideration is "one, many, and unlimited" (16d6). If anyone is to acquire an adequate understanding of pleasure, Socrates wants to say, then he must take account of its unity as well as its diversity. Since Plato is clearly interested in achieving knowledge of pleasure in the *Philebus*, he is obliged—by his own lights—to develop an account of that generic feature which all pleasures share. As Socrates puts it, "whatever is dispersed and split up into a multitude, we must try to work out its unifying nature as far as we can" (25a2-4).[13]

[13] Cf. Irwin (1995, 322-323).

Contrary to the no-theory view, then, it seems plain that Plato has no intention of claiming that pleasures are radically diverse. But then why is this section of the *Philebus* devoted almost exclusively to emphasizing their differences? One apparent reason is that Plato thinks the diversity of pleasure bears directly on the debate about its value. Since Protarchus holds that all pleasures are good (13a8; cf. 13b6-c2) and Socrates holds that only some are (13b1), their dispute obviously cannot get off the ground until Protarchus is persuaded that it is at least possible for pleasures to bear opposite properties (such as goodness and badness). Once Protarchus accepts this point, then he has to provide an *argument* to the effect that all pleasures are good in virtue of being pleasures; he cannot just insist that if goodness is a property of *some* pleasures, then goodness must be a property of *all* of them. So in order to open up the debate, Socrates needs to press the very weak point that it is possible, at least in principle, for pleasures to bear opposite properties.[14]

If this is Socrates' point, then Protarchus resists it not because he thinks that all pleasures have something in common, but because he thinks that pleasures cannot bear opposite properties without ceasing to be pleasures. This would explain why Socrates

[14] Protarchus is reluctant to accept this point, perhaps because he is worried that doing so will fatally compromise his hedonism. If this is his worry, however, then it is unfounded. No division of pleasure alone can threaten the thesis that Protarchus is defending, since any such division merely tells us what all the various sorts of pleasure are, and makes no additional pronouncement about their value. So no matter how the class of pleasure is ultimately carved up, it is still possible (and indeed necessary) for Protarchus to show that every one is good insofar as it is a pleasure. To communicate this point, Socrates concedes that his own intellectualist thesis is in the same position (13e4-10): knowledge too comes in many forms, some of which are unlike each other (ἀνόμοιοί τινες αὐτῶν ἀλλήλαις) and maybe even opposed (ἐναντίαι, 13e10) to each other. While this leaves open the possibility that only some (if any) of these types are good, it does not *foreclose* the possibility that *all* of these types are good. Because hedonism and intellectualism are on the same footing in this regard, Socrates is able to show that neither he nor Protarchus has anything to fear from admitting that their respective candidates for the good take different

moves directly to the examples of color and shape, since colors and shapes obviously can bear opposite properties without ceasing to be colors or shapes. It would also explain why Socrates warns Protarchus against believing "that account which would make all the most opposed things one" (13a3-4; cf. 13c6-d1, 13e9-14a3). On this reading of the passage, Socrates' warning is not directed against those who hold that all pleasures have something in common; it is directed against those who exaggerate the unity of pleasure by refusing to accept that pleasures can be opposed. In short, Socrates wants to wean Protarchus off the idea that if any pleasure is good, then no pleasure is bad. So when Socrates insists that opposed pleasures are not "like one another" (ὁμοίας ἀλλήλαις, 12d4-6), he does not mean (3a) that opposed types of pleasure are not alike in *any* way; he means (3b) that opposed types of pleasure are not alike in *every* way.[15]

One might object that (3b) does not fit the language of (3) nearly as well as (3a) does. After all, to deny that two things are *like* each other is often just to assert that they are completely *unlike* each other. But the Greek does not warrant such a strong reading of (3). Sometimes the force of ὅμοιος is such that there is no real distinction in meaning between ἀνόμοιος and οὐκ ὅμοιος.[16] In such cases, saying that two things are not "like each other" is roughly equivalent to saying that they are not in *every* way alike; it is not equivalent to saying that they are in *no* way alike. If Socrates is using the term in this weaker sense, then in (3) he is asserting only that the "opposed pleasures" are dissimilar;

---

forms. Once Protarchus grasps this, he is able to overcome his initial worries and concede that pleasure (like intelligence) is multiple and diverse (14a6-9).

[15] For a more general but broadly similar take on this issue, see Frede (1993, xviii-xx).

[16] See LSJ ὅμοιος (2), according to which the force of ὅμοιος can be equivalent to the force of ὁ αὐτος.

he is not *also* asserting that they are *wholly* dissimilar. (3b) has just as much philological authority as (3a), and, as we have seen, far greater philosophical authority.

Another possible objection to (3b) as a reading of (3) is that (3b) is redundant. Since Socrates already claims in (1) that the various types of enjoyment are "in some way unlike one another" (τινα τρόπον ἀνομοίους ἀλλήλαις), there is no need for him to add that opposed types of pleasure are unlike each other. For (3b) seems to follow from (1). Now if the passage is best understood as a set of distinct claims about the dissimilarity of various pleasures, then this objection has some bite. But perhaps the passage is better understood as an argument supporting (1). On this reading, Socrates is arguing from (2) and (3b) to (1): since pleasures are taken in different (and even opposite) things, and since pleasures differ from each other insofar as they have different objects, it follows that some pleasures differ from (and are even opposed to) each other. This reading of the passage finds a plausible, non-redundant role for (3b) in Socrates' reasoning, and so saves (3b) as an interpretation of (3). There is no solid evidence in this passage that would commit Plato to the claim that some pleasures have nothing in common. And, as we have seen, there is plenty of evidence to suggest the opposite. The no-theory view, then, is unsustainable.

1.3. Plato's Incomplete Project

Despite his self-acknowledged obligation to provide a unified account of pleasure and pain in the *Philebus*, Plato never explicitly meets this obligation. He provides only a rough sketch of such an account, in the form of his *equilibrium model* (EM). I call this a "rough sketch" because Plato never develops the model in a rigorous way, nor does he clearly show how EM is supposed to apply to any of the various types of pleasure and pain that he discusses. Even if he does think that a substantive theory of pleasure and pain is within reach, he never spells out in explicit detail how the theory is supposed to work. Advocates of the no-theory view argue, based on this apparent neglect, that if Plato had thought that EM could serve as a substantive theory of pleasure and pain, then he would have developed it more openly and effectively than he in fact does.[17]

But there are other possible explanations of Plato's reticence in this regard. It is not implausible to suppose, for example, that he has other reasons for leaving at least some of the theoretical work undone. If the *Philebus* is designed at least in part to provoke directed philosophical inquiry among Plato's Academic students and colleagues, and if Plato supposes that meeting this pedagogical goal requires him to avoid transmitting a specific doctrine that could be learned by rote, then he has a strong reason *not* to spell out his theory in detail.[18] He might also have more straightforward philosophical reasons. His confidence in EM might be shaky, such that he is hesitant to extend it with any

[17] Gosling and Taylor (1982, 140).

precision to all the types of pleasure he recognizes. In that case, he is leaving it up to the philosophically inclined reader to develop EM in a satisfactory way or—if this cannot be done—to abandon it in favor of a different theory. Or perhaps his reasons are both pedagogical and philosophical. In any case, charitable interpreters should take note of Plato's self-imposed obligation to provide a substantive theory, and should do their best to construct such a theory from the materials that Plato makes available in the dialogue.

1.4. The Equilibrium Model

In the course of the *Philebus*, Plato discusses five different kinds of pleasure and three different kinds of pain. I will argue that EM can be applied successfully to each one of them. They are, in order of appearance: (1) bodily pleasure and pain (e.g., eating when hungry, 31e3-32b8); (2) anticipatory pleasure and pain (e.g., expecting, when hungry, that one will soon eat, 32b9-c5); (3) mixed emotional pleasure, when one is pleased and pained simultaneously (e.g., laughing at the failures of one's friends, 47d5-50c3); (4) pure pleasure, when one is pleased without being in pain and without having just been in pain (e.g., listening to a melody or learning something, 50e5-52b8); and (5) virtuous or healthy pleasure (e.g., exercising temperance, 63e4-7; cf. 12c8-d4). I agree with most commentators that EM is designed first and foremost to account for pleasures and pains

[18] For some discussion of Plato's style of writing, and how that style should affect our interpretation of his works, see the essays in Griswold (1988).

of type (1).[19] But I do not believe that EM applies to bodily pleasures and pains in a straightforward way. One might suppose that bodily pleasures, according to EM, are just processes by which some optimal bodily state is restored. On this view, the restoration of an animal body's proper state of moisture is itself a pleasure. I will argue, however, that Plato does not employ EM in this fashion. On my reading, Plato holds that bodily pleasure itself is not a *physiological* process of restoration, but a *psychological* one. In developing this interpretation, I will attempt to show that Plato's more sophisticated use of EM to explain type (1) pleasures and pains reveals a way in which he can use EM to explain types (2)-(5) as well.

At a relatively early stage of the dialogue, Plato advances EM as a way to explain an animal's pleasant or painful experiences. This explanation appeals to processes in which an animal's equilibrium states are destroyed or restored, respectively. Socrates' first stab at this idea conveys what I take to be Plato's main point:

> Now I say that when the harmony dissolves in us animals there comes to be at that time a dissolution of our nature and a coming-to-be of pain [τῆς ἁρμονίας μὲν λυομένης ἡμῖν ἐν τοῖς ζῴοις ἅμα λύσιν τῆς φύσεως καὶ γένεσιν ἀλγηδόνων ἐν τῷ τότε γίγνεσθαι χρόνῳ] … and when [the animal] is harmonized and again returns to its nature, pleasure comes to be [πάλιν δὲ ἁρμοττομένης τε καὶ εἰς τὴν αὑτῆς φύσιν ἀπιούσης ἡδονὴν γίγνεσθαι]. (31d4-10)

When Protarchus seeks clarification, Socrates spells out more precisely what he means by the "harmony in animals" (31c2-d2). He appeals to his earlier "fourfold

[19] See, e.g., Hackforth (1972, 58-59), Gosling and Taylor (1982, 133), Carone (2000, 264-270), and Tuozzo (1996, 495-498).

division of all things" into (1) unlimited things (24a6-25a4); (2) limits (25a6-b3; 25d11-e2); (3) things that result from imposing limits on unlimited things (25e3-26b7; 26d7-9); and (4) things that impose limits on unlimited things (26e1-27c1). Socrates claims that the "harmony" in animals belongs to the third, "mixed" class (31c8-11). Other items in this class, he says, are health (the proper combination of hot and cold, wet and dry in the body, 25e7-8); various harmonic and rhythmic systems in music (the proper combination of high and low tones or fast and slow tempos, 26a2-4); seasonal fluctuations (the proper balance of hot and cold, wet and dry in the atmosphere, 26a6-b3); bodily beauty and strength (26b5-6); and unspecified types of psychological excellence (26b6-7). In each of these cases, Socrates claims, there is a "limit" such as a numerical proportion or ratio (25a6-b3) imposed such that it "prevents the opposites from being at odds with each other, and makes them proportionate and harmonious by establishing number in them" (παύει πρὸς ἄλληλα τἀναντία διαφόρως ἔχοντα σύμμετρα δὲ καὶ σύμφωνα ἐνθεῖσα ἀριθμὸν ἀπεργάζεται, 25d11-e2).

The precise interpretation of these passages is controversial, of course, but their sense is clear enough for my purposes here.[20] In Plato's view, some states of a human agent are good states for her to be in, and their goodness is defined by their being states of *equilibrium*—where this implies that an agent who achieves all of her equilibrium states experiences no conflict between the various opposing forces internal to her. Such states are impermanent, however—at least in mortal beings—and when an agent falls out of one,

she is pained. Likewise, when she is achieving one, she is pleased.[21] Or so Socrates' first formulation suggests.[22]

1.5. The Identity Thesis and Crude Physicalism

The problem with Socrates' first formulation of EM is that it fails to tell us what pains, for example, *are*. All it tells us is that if an animal undergoes a destructive process of some sort, then the animal is in pain. But to specify a sufficient condition for pain is not to specify what pain is. Perhaps in recognition of this shortcoming, Socrates quickly suggests that undergoing a destructive process is sufficient for being in pain because destructive processes *just are* pains. He claims that hunger is "a disintegration [λύσις] and a pain" (31e6), and that eating (when hungry) is "a replenishment [πλήρωσις] and a pleasure" (31e8). Likewise, he claims, thirst is "a destruction [φθορὰ] and a pain" while drinking (when thirsty) is a replenishment (πληροῦσα) and a pleasure (31e10-32a1). Socrates extends this analysis to more complicated cases involving the dissolution and

---

[20] For useful discussions of the fourfold division, see Hackforth (1972, 37-43), Gosling (1975, 153-181 and 185-208), Sayre (1983, ch. 3), Benitez (1989, 59-91), Davidson (1990, 154-325), Frede (1993, xxxiii-xxxix), Irwin (1995, 325-327), and Cooper (1999).

[21] This is broadly congruent with Plato's earlier attempts to theorize about pleasure and pain. See *Gorgias* 492e-499c, *Republic* 583b-586c, and *Timaeus* 64a-65b. For a useful review of these earlier attempts, and an attempt to connect them to the theory developed in the *Philebus*, see Frede (1992, 433-437).

[22] It might be thought that this formulation is at variance with what Socrates says earlier about pleasure's place in the fourfold division. Before introducing EM, Socrates follows Philebus' suggestion and assigns pleasure to the unlimited class (28a3-4). "Pleasure itself," he says, "is unlimited and belongs to the class that, in itself and by itself, neither has nor ever will have beginning, middle, or end" (μήτε ἀρχὴν μήτε μέσα μήτε τέλος ἐν αὑτῷ ἀφ' ἑαυτοῦ, 31a7-10). But this claim seems to drop out of the picture when Socrates introduces and develops EM. According to EM, pleasure is a process in which items in the mixed class are restored—and the items in the mixed class are themselves combinations of limits and unlimiteds. So one might wonder whether Plato is entirely consistent on this point. I suspect that he is. After all, there is no contradiction in claiming that restorations of things in the mixed class belong to the unlimited class.

coagulation of the flesh via heating and cooling. In these cases too the basic idea is the same: when the process is "against nature" (παρὰ φύσιν) it is a pain (32a3; a6-7); when it is "according to nature" (κατὰ φύσιν) it is a pleasure (32a3-4; a7-8). Socrates sums up his position (32a8-9) as follows: "whenever [the harmony in organisms] is destroyed, the destruction is pain [τὴν μὲν φθορὰν λύπην εἶναι]; but as for the path to their being, this returning restoration of them all is pleasure" (τὴν δ' εἰς τὴν αὐτῶν οὐσίαν ὁδόν ταύτην δὲ αὖ πάλιν τὴν ἀναχώρησιν πάντων ἡδονήν, 32b2-4). According to this new formulation, then, restorative and destructive processes in animals are not only sufficient for pleasures and pains; they *are* pleasures and pains.[23] This amounts to a version of EM that specifies more precisely what bodily pleasures and pains are. I will call it the *identity thesis* (ID). According to ID, all bodily pleasures and pains are identical to restorative and destructive animal processes, and vice versa.[24]

The problem with ID, at least as it stands, is that it has strange and implausible consequences. Remember that, according to Socrates, many of the processes to be identified with pleasures and pains are physiological. "Congealing" of the flesh (32a7) and "replenishment of the dry [body] with liquid" (31e10-32a1) are pretty clearly physiological processes, yet according to Socrates they are pains and pleasures

---

Plato might assign pleasure to this class simply because pleasures exhibit indeterminate fluctuations in intensity, duration, and size. See Cooper (1999).

[23] Tuozzo (1996, 499-500) acknowledges that this is the thrust of Socrates' remarks here, but suggests that Socrates makes this point for purely pedagogical reasons and should not be interpreted as having genuinely endorsed it. When faced with 42c9-d8—a later passage in which Socrates restates the same position—Tuozzo submits that he is merely setting the view up for refutation. This reading is obviously strained. Moreover it is motivated by Tuozzo's assumption that Socrates ultimately rejects the idea that pains are destructions. As we shall see, this assumption is dubious.

[24] As I understand it, ID is not limited to the claim that all destructive processes are bodily pains. Since ID is supposed to provide an explanation of what bodily pains are, the converse is implied.

respectively. If this is what Plato means to say here, then he is endorsing a view on which the pain of burning one's hand, say, *just is* the rapid or intense overheating of one's hand. I will call this view *crude physicalism*.

Crude physicalism, whether Plato holds it or not, is vulnerable to a serious objection. For whatever else they may be, pains are—or at least involve—states of consciousness. Crude physicalists have to deny this, since the overheating of a hand neither is nor presupposes any state of consciousness in the agent whose hand it is. So the pain of burning one's hand, on their view, cannot be relieved unless and until one's hand returns to its equilibrium temperature. Faced with the success of analgesics, the crude physicalist has to force a distinction between pain and the *feeling of* pain—such that it is possible to have the former without the latter. Pains can be unfelt, on this view, because they are not themselves feelings. Yet at this point it seems that the crude physicalist, in order to remain coherent, has to distort the core concept of pain rather badly. It is hard to understand, especially in the context of physical pain, what might be meant by speaking of "unfelt pain" or "being in pain but not feeling it." So if we find that Plato holds this view, then I think we should conclude that he is mistaken.

But evidently Plato does not hold this view. In a later section of the dialogue (42c-44b) he has Socrates advance an argument to the effect that undergoing a destructive bodily process is not even sufficient for being in pain. The broader goal of this argument is to show that an utterly apathetic life is possible, but not necessarily pleasant—since being free from pain is not itself a pleasure (44a4-10). Socrates leads off the argument by

restating ID (42c9-d7), but soon makes three related claims that together constitute a serious challenge to it:

(1) Animals who undergo bodily processes frequently do not perceive them (ταῦτ' αἰσθάνεται). (43b1-6)

(2) If an animal does not perceive the destructive bodily process that it is undergoing, then the animal does not undergo the pain connected with that process. (43b7-10)

(3) If a destructive bodily process has enough "size" (μεγάλαι) to be perceived by the animal, then the animal does undergo the pain connected with that process. (43c4-6)

Socrates suggests that (2) and (3) require modifications of previous agreements (43b7-c2), and it's not hard to see why. If an animal can undergo a destructive bodily process without being in pain, then ID—strictly interpreted—is false. For some destructive processes are not pains. When destructive bodily processes do not have enough "size," they fail to register in the animal's soul; and when they fail to register in the animal's soul, the animal is not in pain. As Socrates puts the same point in an even earlier discussion, it is possible for bodily processes to be "extinguished in the body before they get through to the soul, leaving [the soul] unaffected" (τὰ μὲν ἐν τῷ σώματι κατασβεννύμενα πρὶν ἐπὶ τὴν ψυχὴν διεξελθεῖν ἀπαθῆ ἐκείνην ἐάσαντα, 33d2-4; cf. 33d8-10). Since Plato clearly concedes that animals can undergo such bodily processes without being in pain, he must reject crude physicalism.[25]

[25] This is widely recognized by commentators. See Gosling and Taylor (1982, 178-180) for an especially careful discussion.

1.6. Perceptual Physicalism

The most obvious way for Plato to reject crude physicalism without losing hold of EM is to adjust ID so as to rule out the possibility that unperceived bodily destructions might count as pains. He can do this by stipulating that destructive bodily processes are identical to bodily pains *just in case these processes are perceived by the animal undergoing them*. On this new version of the thesis, not all destructive bodily processes are pains. Only (and all) those that are *perceived* are pains. Consequently I will call this view *perceptual physicalism*.[26]

One reason to resist attributing this view to Plato, however, is that it has consequences that would be philosophically unappealing to him. According to the perceptual physicalist, one of the pain-making properties of a destructive bodily process is relational. That is, what makes this bodily process a pain is, at least in part, its being an object of awareness; but "being an object of awareness" is not an intrinsic property of any bodily process. Imagine, for the sake of illustration, two type-identical bodily destructions in two type-identical animals—one alert, the other anaesthetized. Even though (by hypothesis) these two bodily destructions share all of their physical

[26] If I understand them correctly, both Frede (1992, 441, 444) and Van Riel (2000, 22-27, 41) attribute this view to Plato. Frede says, somewhat strangely, that "pleasures and pains are never a matter of the body alone, because only those disturbances in the body are pains and those restorations pleasures that are perceived by the soul. This leads to the definition of the so-called pleasures of the body as perceptions." But since *perceived processes* are obviously not equivalent to *perceptions of processes*, it is not at all clear what she is trying to say here. Later (444) her claim seems to be that pleasure is defined "as a perceived filling or restoration," which I take to be a straightforward formulation of perceptual physicalism. Van Riel holds that pleasure "always is the replenishment of a lack, with the additional qualification that the replenishment should be perceived in order to yield pleasure" (27). Given that it is simply incoherent to

properties, the perceptual physicalist has to claim that only the first one is a pain. And this, I take it, is tantamount to claiming that a pain's definitional properties can be conferred by mere Cambridge changes. For if the second animal were revived as a result of some sudden contingency, then the destructive process in that animal would at that moment become a pain. But then it seems that part of what makes a destructive bodily process a pain is a property that is accidental to it. This would almost certainly be unacceptable to Plato, since he presumably thinks that the real definition of a thing picks out only its essential properties. Consequently he cannot—or at least should not—accept perceptual physicalism.

Another reason to resist attributing this view to Plato is that the thrust of his argument here strongly suggests that he rejects it. When Socrates claims that a destructive process is sufficient for bodily pain only if it has enough "size" to register in the psyche, the language he uses is striking: he says that only "sizeable" bodily processes "accomplish" (ἀπεργάζονται, 43b9) or "produce" (ποιοῦσιν, 43c5) pleasures and pains. Since both of these verbs have clear causal force, the passage as a whole implies that bodily pains are *effects* of destructive bodily processes. But this is plainly incompatible with the claim that bodily pains *just are* such processes, even if it is conceded that these processes must be perceived.[27] So if Plato holds to his language in this passage, he must reject perceptual physicalism as well.

---

claim that a pleasure is something that yields pleasure when it is perceived, Van Riel's position here is either perceptual physicalism or nothing.

1.7. The Causal Theory

Given that Plato must reject both crude and perceptual physicalism, it might seem that he must also reject any version of the identity thesis, no matter how it is weakened or refined. For if destructive bodily processes *cause* bodily pains, then clearly they cannot be *identical* to bodily pains. Recognizing this problem, Tuozzo (1996) argues that Plato ultimately abandons the identity thesis in favor of the theory that bodily pains are those conscious mental processes that are caused by destructive bodily processes. I will call this view the *causal theory*.

As we shall see, the causal theory holds great promise. But it is also beset with problems, both as an interpretation of Plato's view and as a free-standing account. If Plato were to endorse such a theory, then he would be committed to the claim that one can reach understanding of what bodily pleasures and pains are by grasping their antecedent causal conditions. A claim of this sort has two crucial components, I take it: first, that there is nothing more to bodily pleasures and pains than their causal roles; and second, that their causal roles are given entirely by their causal antecedents. Thus the causal theory is a peculiar hybrid of functionalism and epiphenomenalism, according to which *what it is to be* a bodily pain is exhausted by whatever causes it, and whatever it causes is strictly irrelevant to what it itself is. A corollary of this view, as Tuozzo (1996, 505) himself seems to recognize, is that bodily pleasures and pains as such have no essential effects on the behavior of animals who undergo them—for otherwise these

[27] Cf. Tuozzo (1996, 502).

effects would have to be taken into account by the theory.[28] So if Tuozzo is right, then Plato must think of bodily pleasure as merely "a psychic epiphenomenon of the restoration that causes it" (Tuozzo 1996, 497).

But Plato would almost certainly deny this. In his lengthy discussion of desire (34d-35d), Socrates claims that bodily pains such as hunger and thirst are also "appetites" (ἐπιθυμίας, 34d10-e14) and that each of them is an "impulse" (ὁρμή, 35c12) in the animal such that it now "strives" (μηνύει, 35c10) to achieve the object of its appetite. He then claims that the objects of appetite are fixed in every case by what the animal takes to be "the pleasant things that would put a stop to the pain" (τῶν ἡδέων ... παύοιτ' ἄν τῆς ἀλγηδόνος, 35e9-10). These passages strongly suggest that Plato, in his account of animal behavior, assigns explanatory roles to bodily pleasure and pain that go well beyond their being caused by restorative and destructive bodily processes. If so, then Plato could not accept the epiphenomenalist position Tuozzo attributes to him, nor could he accept the causal theory from which it follows.

Moreover, Plato has good independent reasons to reject the causal theory. These reasons emerge most clearly, I think, from a close reading of Plato's brief remarks on perception. When he has Socrates raise the idea that a bodily process might escape awareness, Socrates gives an interesting description of what happens when human beings perceive:

[28] Of course the causal theorist can claim that bodily pleasures and pains have effects. What the causal theorist cannot admit is that these effects are essential to their causes. If this were admitted, then there is more to what bodily pleasures and pains are than what causes them.

> Some of our bodily affections (παθημάτων) are extinguished in the body before they get through to the soul, leaving it unaffected (ἀπαθῆ). Others go through both and inflict a kind of disturbance that is peculiar and common to each (τινα ὥσπερ σεισμὸν ἐντιθέντα ἴδιόν τε καὶ κοινὸν ἑκατέρῳ). (33d2-6)

According to Socrates, perception is what happens "when the soul and body are moved in common by a single affection" (ἑνὶ πάθει τὴν ψυχὴν καὶ τὸ σῶμα κοινῇ ... κινεῖσθαι, 34a3-5). Indeed, the perception *just is* the single motion (κίνησιν) that affects them both (34a4). Now destructive bodily processes are no doubt among those "affections" or "disturbances" in the body that can reach, or fail to reach, the soul. If they fail to reach it, then presumably they cause no pain; but if they succeed in reaching it—and perception occurs—then presumably they do cause pain.

Now consider a hypothetical case in which an animal's soul is not moved by any destructive bodily process, but for some reason *is* moved in just the way it *would* be moved if it *were* moved by a destructive bodily process. Clearly this is a case of illusion or hallucination, not perception. But the crucial point is that it also seems to be a case of *pain*. For if the animal's soul were affected as it would be by a strong destruction of its body, and if the animal's pain-experience is explained by the affections of its soul, then evidently we should say (on Plato's behalf) that pain does occur—even if there is no bodily destruction causing it. Indeed, we should say that this illusory pain is indistinguishable (on the "inside," as it were) from the pain that the animal *would* undergo if it *were* undergoing a destructive bodily process. And this, I take it, is just to say that bodily pains in general neither are nor require perceptions. Rather, they are disturbances

of the soul that can be—but need not be—caused by disturbances of the body. The animal might suffer what are commonly called "phantom pains"—bodily pains that are *as if of* bodily destructions but not *caused by* bodily destructions. Since Plato himself, in his discussion of "false" pleasures and pains, indicates that one can be pained by "what is not" (40d7-10), he has grounds for denying that bodily pains must be caused by destructive bodily processes.

1.8. The Psychological Identity Thesis

Despite its flaws, the causal theory is promising. For it holds that bodily pains are identical to *psychological* processes rather than *physical* ones. If Plato accepts this part of the view—and rejects the quasi-functionalist, epiphenomenalist part—then he has a potentially fruitful way to modify the identity thesis so as to address the various problems he himself raises. These problems, remember, make it impossible for him to accept either crude or perceptual physicalism. This in turn makes it difficult for him to sustain the identity thesis (as interpreted under EM), which is—for better or worse—the only viable theory that Plato puts on offer in the entire dialogue. But he can save this thesis after all if he denies that bodily pains are identical to destructive bodily processes and affirms that bodily pains are identical to destructive *psychological* processes. This amounts to a narrowed-down version of the identity thesis. On this view, bodily pains

are still identical to destructions, though these destructions are now specified as psychological. So I will call this the *psychological identity thesis* (PID).

Some commentators suggest that Plato, in his analysis of pleasure and pain, is not interested in applying EM to anything other than bodily states or events.[29] If so, then Plato cannot hold PID, since PID depends crucially on the idea that there are psychological restorations and destructions. One reason to go along with this is that almost all of the examples Socrates provides in his discussion of EM and ID seem to be bodily processes. He claims that the "congealing" of the body (32a7) is both a destructive process and a pain, and that the "replenishment of the dry [body] with liquid" (31e10-32a1) is both a restorative process and a pleasure. These two processes certainly seem to be bodily, as do some of the other processes he mentions, such as eating (31e8).[30] But it is important to notice that Socrates, in the same passage, applies EM to hunger (πείνη, 31e6) and thirst (δίψος, 31e10)—even though he seems not to think of these two processes as bodily. Indeed, he claims emphatically that "our argument will in no way accept that the body undergoes thirst (διψῆν) or hunger (πεινῆν) or anything of the sort" (35d5-6). Since there is no indication that Socrates has at this point given up his claim that thirst and hunger are destructive processes, there is good reason to suppose that he thinks of thirst and hunger as destructive *psychological* processes. Moreover, in his much

[29] Gosling and Taylor (1982, 132-136) are the most prominent defenders of this view.
[30] It is worth keeping in mind, however, that Socrates might be deploying these examples only because he thinks they will be easier for Protarchus to grasp on first hearing. Both Socrates and Protarchus acknowledge that the initial formulation of EM is unclear (31d9-e2), and this prompts Socrates to illuminate it by mentioning a set of what he calls cases that are "easiest to understand" (31e3-4). Given these remarks, we should perhaps be wary of assuming too much about EM based on the examples that

later discussion of the so-called "pure pleasures," Socrates claims unambiguously that the pleasures of learning are ways of "being filled" (πληρωθεῖσιν, 51e7-52a3), and this implies that the pleasures of learning are restorative psychological processes. None of this is particularly surprising, I take it, given that Socrates includes on his list of harmonious states an unspecified but wide-ranging set of optimal psychological states (ἐν ψυχαῖς αὖ πάμπολλα ἕτερα καὶ πάγκαλα, 26b6-7).[31] Since every other harmonious state on this list is apt to be destroyed and restored, there is no compelling reason to suppose that the scope of EM is restricted to the body. In sum, Plato seems perfectly comfortable with the idea of psychological restorations and destructions. Hence there is nothing in Plato's deployment of EM alone that prevents him from holding PID.

Some positive indications that Plato holds PID will emerge, I think, from a careful examination of his argument against crude physicalism. That argument, remember, establishes that an animal's perceptual awareness (αἴσθησις, 43b2) of a destructive bodily process is sufficient for that animal to be in bodily pain, and that a destructive bodily process, if it is to "accomplish" or "produce" pain, must reach the animal's psyche. The question to ask, given this framework, is whether the animal's pain is (i) an exclusively bodily process, (ii) an exclusively psychological process, or (iii) a process that is both bodily and psychological. Since Plato thinks that perceptual awareness is at least partly psychological (e.g., 34a3-6), he must reject (i). For the only way he can accept (i)

---

follow. Perhaps these "commonplace and obvious" cases are deployed precisely in order to illuminate more difficult cases, such as psychological restorations and destructions.

[31] For an extended and wide-ranging defense of the idea that Plato is deeply committed to the idea that the soul has an optimal equilibrium state, see Gocer (1999, esp. 24-33).

and still maintain that perception is partly psychological is for him to claim that bodily pains are just destructive bodily processes that are perceived. But to make this claim is to endorse perceptual physicalism, and—as I have already argued—Plato rejects that doctrine.

Nor does Plato seem to think that bodily pains are psycho-physical processes. In his brief but interesting discussion of perception (αἴσθησις, 33c-34a), Socrates provides a general description of what happens when human beings perceive. He claims that when we have perceptions, some physical disturbance works its way through the body to the soul, inflicting a "disturbance that is peculiar and common to each" (i.e., the body and soul) (σεισμὸν ἐντιθέντα ἴδιόν τε καὶ κοινὸν ἑκατέρῳ, 33d2-6). Whenever these bodily disturbances fail to reach the soul, he says, the soul is "oblivious" (λανθάνειν) of them (33d8-10). This state of oblivion he calls "non-perception" or "anaesthesia" (ἀναισθησίαν, 33e10-34a1). Perception (αἴσθησις), on the other hand, happens "when the soul and body are moved in common by a single affection" (ἑνὶ πάθει τὴν ψυχὴν καὶ τὸ σῶμα κοινῇ ... κινεῖσθαι, 34a3-5). Indeed, the perception *just is* the single motion (κίνησιν) that affects them both (34a4). Now these somewhat cryptic remarks may not settle by themselves the question of whether bodily pain is a psycho-physical event or an exclusively psychological event, but they do help to frame the debate. Since Plato implicitly refers back to these remarks in his argument against crude physicalism, he seems to think that the model of perception on offer here is relevant for understanding what—and where—bodily pain is. The debate, then, can be framed roughly as follows.

Given Plato's model of perception, what is it that bodily disturbances cause when perception occurs? Do they cause exclusively psychological disturbances? Or do they cause disturbances that are somehow both psychological and physical?

The only clear reason to attribute to Plato the claim that bodily pains are psycho-physical (rather than exclusively psychological) is that he takes *perceptions* to be psycho-physical. On his view, remember, a perception is a motion that affects the body and the soul at once. So if we suppose that Plato thinks of bodily pains themselves as perceptions of a certain kind, then we might be tempted to conclude that these pains are psycho-physical in nature. But it would be a mistake to draw this conclusion, I think. For Plato never actually suggests that bodily pains are perceptions. In all the relevant passages, Socrates says only that perception of a destructive bodily process is sufficient for pain-experience; nowhere does he claim that this is so precisely because bodily pains themselves are perceptions of destructive bodily processes. Nor should he claim this, given his previous commitments. For if destructive bodily processes are supposed to cause bodily pains when they "reach" the soul, then it is very difficult to see how pains could be perceptions—at least as Plato understands them. For a case of perception, on his account, is a case in which a single motion agitates both the body and the soul. Indeed, that single motion *just is* the perception. But then a motion in the body that is strong enough to move the soul is not the *cause* of a perception; it *is* a perception. (The cause of a perception, presumably, is some state or event in the world external to the perceiver.) So if Plato holds that bodily pains are the effects of bodily processes, then he cannot hold

that these pains themselves are perceptions.[32] The most he can say is that the perception of a destructive bodily process ensures the occurrence of bodily pain.

In fact there are good independent reasons for him not to say anything more than this. Consider the case of "phantom pain" that I raised earlier. In such a case the animal's soul is not moved by any destructive bodily process, but nonetheless is moved in just the way it *would* be moved if it *were* moved by a destructive bodily process. Though this is not a case of perception, it certainly seems to be a case of pain.[33] If so, then bodily pains neither are nor require perceptions. Now if we switch our attention for a moment from phantom pain to phantom pleasure, we find Socrates endorsing something close to what Plato needs here: "whoever has any pleasure at all … really does have pleasure, even if it is not about anything … that is the case" (40d). On a liberal reading of this passage, Socrates is claiming here that phantom pleasures really are pleasures, even though they are not caused by any restorative bodily process. And this suggests that, on Plato's view, bodily pleasures and pains are themselves purely psychological—despite their (normally) having physical causes.[34]

---

[32] *Contra* Gosling and Taylor (1982, 180-183) and at least one reading of Frede (1992, 441).
[33] It is not a case of perception in Plato's sense of that term, since he implies (at 34a) that perception always involves an accurate registering of bodily motion in the soul.
[34] As Hackforth (1972, 61 and 112) points out, Socrates implies as much when he challenges the hedonist to make sense of the idea that "there is nothing good or noble in bodies or in anything else except the soul, and even there pleasure is the only good thing" (55b1-3). The assumption that pleasure and pain are exclusively psychological events seems to be one that is shared by both Socrates and his opponent. Cf. Davidson (1990, 335).

1.9. Intentionalism

As we have seen, Plato seems to think that bodily pains are exclusively psychological events, and that EM can in principle be applied to psychological as well as physical events. But it remains unclear how and why EM is supposed to apply to a bodily pain when that pain is understood as an exclusively psychological event. Plato's discussion of perception, once again, can help clear things up. In the case of perception, Socrates claims, "the soul and the body are moved in common by a single affection" (34a3-5). Now it might be thought that the "single affection" that disturbs the soul in this case is itself a *bodily* disturbance, since this is precisely what is said to move the soul when perception occurs. But if the bodily affection and the psychological affection are supposed to be identical in such cases, then Plato's account generates some unfortunate consequences. Suppose, for example, that the burning of an animal's limb—a destructive bodily process—is intense enough to reach the animal's soul and move it. Now if Plato holds that the motion in this animal's soul is identical to the one in its body, then he has to claim that in this case the *animal's soul itself* is burning—which is absurd, especially by Plato's lights. Plato never seriously entertains the idea that the soul is a body, and so would presumably deny that souls can be affected by bodily changes in the same way that bodies are. So when Plato claims that the body and the soul are moved by a "single affection" in cases of perception, he cannot mean that this affection is uniform.

Indeed, in an earlier passage Socrates suggests that the overall picture is more complicated than this. Before giving his definition of perception, Socrates claims that an animal perceives just in case its body and soul together undergo a disturbance that is "both peculiar to each and common to each" (ἰδιόν τε καὶ κοινὸν ἑκατέρῳ, 33d5-6). This formulation, unlike the one we have just examined, implies that in perception the body and the soul are not affected in the same way. There is something "peculiar" about the way the soul is affected by the body, even though this affection is somehow shared by both. But then how is it shared? In my view, the best way to make sense of this paradoxical claim is to say (on Plato's behalf) that in cases of perceptual bodily pain the soul does not undergo the same *physical* change as does the body; it undergoes a peculiarly *psychological* change, albeit one that shares some important structural property with the physical change that causes it. And the only plausible candidate for such a property that ever comes up in the *Philebus* is that of *being a destruction* in the case of pain, and *being a restoration* in the case of pleasure. In the standard case of bodily pain, then, the body undergoes a disruption of physical harmony such that this disruption reaches the soul and triggers a distinct disruption of psychological harmony. The soul undergoes a *peculiar* change insofar as this change is psychological rather than physical; but the soul undergoes the *same* change insofar as this change is a disturbance of some equilibrium state.

For Plato this connection between bodily disturbance and psychological disturbance is not especially mysterious, since in his view the soul is responsible for

nurturing and sustaining the body it controls.[35] He classifies all souls as things that impose limit and order on unlimited things, and by this imposition generate in bodies a set of equilibrium states that are optimal for them. Each soul has the task of establishing and maintaining such states in the body to which it belongs, whether that be "the body of the universe" or "the body that is ours" (29e1-30a7). As Socrates puts it, the soul in us "provides training for the body and healing for its wounds, and in other cases other kinds of order and repair" (30b1-3) just as the soul of the universe "orders and coordinates the years, seasons, and months" (30c5-6). One of the soul's functions, then, is to preserve equilibrium in a body that is helpless to do this on its own. What Plato suggests here, but never explicitly states, is that the soul's own equilibrium state depends at least in part on its satisfactory performance of this ordering function with respect to the body. A soul that recognizes that its body is urgently in need of "restitution" is to this extent a *disturbed* soul—one that is, for the moment at least, knocked out of its *own* equilibrium. If this is Plato's view, then he must assume that an animal's soul undergoes its own peculiar destruction when it registers a destruction in the body it governs. This psychological disturbance would be equivalent to a kind of alarm at the disrupted condition of the body and, at the same time, an impulse to repair it.

But if the soul is to perform its governing role properly, the connection between bodily and psychological disruption needs to be tighter. In the case of perceptual pain, it is not enough for a bodily disturbance to cause any old psychological disturbance. The soul's disturbance has to carry some sort of *information* about the body's disturbance;

[35] See also *Phaedrus* (246b6).

otherwise the soul will be unable to perform whatever tasks are necessary to restore the body's equilibrium. Plato can make room for this requirement by claiming that bodily pains cannot fail to have *body-oriented intentional content*. On this view, all bodily pains must have content—*that the left arm is overheating*, say—even when there is nothing happening in the body that could satisfy this content. By modifying his account in this way, Plato can embrace PID as a theory of bodily pain without compromising any of his other commitments. Bodily pains are still destructions on this view, but they are *psychological* destructions with contents that specify conditions of *bodily* destruction. I will call this view *intentionalism*.

If Plato accepts intentionalism, he can escape several of the problems that he raises for himself. He can retain a fairly robust version of the identity thesis and thereby keep his hope for a substantive theory alive; he can make sense of the soul's ability to promote equilibrium states in the body; and he can accommodate the otherwise troublesome cases of phantom pain and phantom pleasure. He can do all this by making two crucial claims: first, that every bodily pain is identical to a disruption of psychological equilibrium; and second, that these psychological disruptions have a kind of body-oriented content, such that they are intrinsically sensitive to disturbances of equilibrium in the body. With this sort of view in hand, Plato can account for cases of (non-hallucinatory) pain roughly as follows. When one's hand is burning, there is a disturbance of the hand's proper balance between hot and cold. Given that this destructive process is intense, it reaches the soul and causes a distinct destructive process

there. This second, psychological disturbance is not itself a disturbance of a bodily balance between hot and cold; rather, it is a disturbance of what we might call *psychological equanimity with respect to* a bodily balance between hot and cold. The soul is *upset*, as it were, *about* the hand's burning. It is in this sense that the soul undergoes a destruction of its own—a destruction of psychological harmony that is tied both causally and intentionally to the state of the hand. And this psychological destruction *just is* the pain of burning one's hand. To generalize, then, each perceptual (non-illusory) bodily pain is identical to some psychological destruction that takes as its intentional object some destruction of an equilibrium state of the body.[36]

All things considered, then, intentionalism is the view that best suits Plato's overall project. It provides the bedrock of a substantive theory of bodily pleasure and pain without violating any of his other commitments. Moreover, I believe there are good independent textual grounds for thinking that he holds it—or at least something very close to it.

[36] The intentionalist need not claim that the pained animal is always aware of the *cause* of its pain, even in perceptual cases. As Tuozzo (1996, 505) notes, the pleasure-taker would have to grasp and apply the concepts of a fairly sophisticated physiological theory (such as EM) in order to recognize a pleasure's cause as such. And it would be odd to suggest that an animal can have bodily pleasure-experience only if it grasps and applies such concepts. But Tuozzo seems to suppose that if a pleasure has any intentional content at all, then the pleasure-taker must be aware of the pleasure's cause. This is incorrect. Surely I can enjoy something about the experience of a steam bath without being aware of the physical facts about steam and skin that causally explain my being pleased. Indeed, I can enjoy something about the bath without even being aware of *what it is about the bath I am enjoying*. It often takes some work on our part to figure out what it is about an activity or an object that pleases us. This is one reason why we have culture critics. Cf. Gosling and Taylor (1982, 182).

1.10. The Intentionality of Desire and Pain

In his discussion of desire and its role in generating animal action (34c-35d), Socrates advances an argument to the effect that "the body never undergoes thirst or hunger or any such thing" (35d5-6). Desire, he concludes, is an exclusively psychological phenomenon (35d1-3). The argument itself is long, tortuous, scattered, and strange; it also relies on several tacit premises. But I believe it provides an invaluable line of entry into Plato's philosophy of mind, and for this reason alone I think it worthwhile to lay out the argument in as complete a way as possible. To that end I offer the following (admittedly cumbersome) reconstruction. Let T be a thirsty man.

(1) T desires filling his body with drink. (34e9-35a1)
(2) T's body is being emptied (of liquid). (35a3-4; 35b3-4)
(3) If a body is being emptied, then it is not being filled. (35a3-4; 35b3-4)
(4) If T desires filling, then there is something that belongs essentially to T which is "in touch with" filling. (35b6-7)
(5) There is something that belongs essentially to T which is "in touch with" filling. (From 1 and 4)
(6) There is something that belongs essentially to T which is the source of the thirsty man's desire. (Tacit)
(7) For any X, X is the source of T's desire only if X is in touch with filling. (Tacit)
(8) A body is in touch with filling only if it is being filled. (Tacit)
(9) T's body is not being filled. (From 2 and 3)
(10) T's body is not in touch with filling. (35b9) (From 8 and 9)
(11) T's body is not the source of T's desire. (35d5-6) (From 7 and 10)
(12) A man's soul and a man's body are the only two things that belong essentially to him. (λοιπόν, 35b11)
(13) T's soul is in touch with filling. (35b11) (From 5, 10, and 12)
(14) T's soul is the source of his desire. (35d1-3) (From 6, 11, and 12)

(15) If a soul is in touch with filling, then it is in touch either through perception or through memory. (35a6-7; cf. 34a10-11)
(16) T's soul is in touch with filling through perception only if T's body is being filled. (34a3-5)
(17) It is not through perception that T's soul is in touch with filling. (35b11-c1) (From 9 and 16)
(18) T's soul is in touch with filling through memory. (35b11-c1; cf. 35d1-2) (From 13, 15, and 17)

Several of the claims Socrates makes in this argument have implications that bear directly on the question of whether Plato endorses intentionalism. In order to see how these implications emerge, however, we must first recall Plato's earlier discussion of bodily pain. There, remember, he claims that "hunger" (πείνη, 31e6) and "thirst" (δίψος, 31e10) and the like are destructions and pains. But here, less than four Stephanus pages later, he claims that "hunger and thirst and many other such things" are *desires* (ἐπιθυμίας, 34d10-e1). Since Plato never indicates that there is some shift in the reference of these terms from one passage to the next, I think it fair to suppose that he takes bodily desires *to be* bodily pains—and *vice versa*.[37] This point is especially significant for our appreciation of the argument, since it licenses us to extend Plato's premises and conclusions about bodily desires to bodily pains as well. Once we do this, two important features of Plato's account of bodily pain will come into view. We will see, first, that Plato thinks of bodily pains as psychological states that have intentional content. Then we will see that he takes these psychological states to be the sources of an animal's motivation to repair the bodily destructions that give rise to them. These two views, taken together, suggest the further view that bodily pains are psychological destructions

that (normally) take bodily destructions as their intentional objects. But before drawing out this further view, I need to establish that Plato's argument underwrites the first two.

The first thing to notice about Plato's argument is that it turns on the idea, introduced in premise (4), that the desiring animal must somehow be "in touch with" (ἐφάπτοιτο) the object of its desire (35a7, b7, b11, c1). It is unclear at first what concept Plato is trying to deploy here, not least because his terminology is so alien to us. But as the argument unfolds it becomes apparent that the concept is meant to capture an important condition for the possibility of desire-based action. In Plato's view, an animal A can experience desire only if it desires some object x (35b1); and for A to desire x is for A to be motivated to get x (35c9-10). Yet in order to be motivated to get x—Plato seems to think—A must have some sort of *contact* with x. For if A has no contact *in any sense* with x, then there can be nothing in A that is (as it were) *fixed upon* x. What Plato is trying to tell us, I think, is that if there is nothing in A that is in some sense fixed upon x, then A cannot be motivated to get x. For in lacking any contact of any sort with x, A could not intelligibly *direct itself at* x—and so could not possibly be *motivated to get* x.

Once this point is established, the rest of the argument unfolds as an attempt to specify more precisely what kind of contact A must have with x, and how this contact can be achieved. In the crucial premise (8), Socrates implies that the contact in question cannot be *physical*—after all, it is not as though A is fixed upon x in the required sense by *physically* attaching to x.[38] The alternative, which Socrates spells out from (15)-(18), is

[37] For further defense of this claim, see Frede (1992, 441) and (1993, 35n2).
[38] Cf. Davidson (1990, 343-345).

that A is fixed upon x in the sense that A *represents* x to itself as something to aim for in action, and does so via the psychological operations of memory. Plato's thought seems to be that A can achieve the required contact with x by remembering x as the sort of thing to be had, given circumstances similar to those that A currently finds itself in. The objects of A's desire are determined, he suggests, by what A "remembers" as "the pleasant things that would put a stop to the pain" (μεμνῆται δὲ τῶν ἡδέων ὧν γενομένων παύοιτ᾽ ἄν τῆς ἀλγηδόνος, 35e9-10). The upshot of his account, then, is that animal motivation itself is possible only if the animal's desires take objects that are represented as pleasant by psychological operations such as memory. So all his strange talk of agents having to be "in touch with" what they desire is not mere bluff. On the contrary, it is the sign of a serious struggle on his part to illuminate what he takes to be the necessary intentional structure of human agency.

According to Plato, then, all bodily desires are bearers of intentional content. However, as we have already seen, he also thinks that all bodily desires are bodily *pains* (and vice versa). He is therefore committed to the claim that all bodily pains are bearers of intentional content. Taken by itself, this claim is perfectly consonant with my interpretation of Plato's account. But it is grounded in the previous claim that bodily pains and bodily desires are identical to each other, and this seems to generate a problem. For if it is assumed that intentional states are individuated solely by their contents, then bodily pains and bodily desires are identical to each other only if they have the same content. Yet it is obvious that they have very different sorts of intentional content. The

content of A's bodily pain is satisfied if there is a destruction of some equilibrium in A's body; the content of A's concomitant desire is satisfied if there is a *restoration* of that same equilibrium. This should be enough to distinguish the pain of thirst, say, from the desire of thirst. But Plato seems oblivious to the distinction.

If pressed, however, he might deny the assumption that mental states are always individuated by their contents. He might urge that there is just one type of mental state that is responsive to bodily damage, even though each state of this type has both an information-bearing role and an action-guiding role. On this view, part of what it is to be responsive to bodily damage is to *register* this damage when it happens; the other part is to *repair* this damage. If this is how Plato is thinking, then perhaps he has good reason to deny that these two roles must be played by different mental states. To take pain in the destruction of a bodily harmony, he might say, is just to desire the restoration of that harmony; and to desire the restoration of a harmony is just to take pain in the destruction of that harmony. For Plato, then, pain and desire might simply be two faces of the same response. He might be wrong about this, of course. But even if he were, his error on this point would not by itself support the view that he is irredeemably confused about the different functional roles that need to be captured by his theory.

1.11. The First Plank of the Theory

Let us take stock. According to the theory I am attributing to Plato, bodily pains have three essential properties: (i) they are destructions of psychological equilibrium; (ii) they have intentional content that, when satisfied, refers to destructions of bodily equilibrium; and (iii) they motivate the animal who undergoes them to repair the bodily destructions that figure in their intentional content. Normally, but not always, a bodily pain is caused by the bodily destructions picked out by the pain's intentional content. Bodily pleasures, on the other hand, have similar essential properties: (i) they are restorations of psychological equilibrium, and (ii) they have intentional contents that, when satisfied, refer to restorations of bodily equilibrium. Instead of being states of *alarm*, however, they are states of *relief*. As a result, they do not have the same motivational force that bodily pains do. Bodily pleasures are signals that bodily harmony is being restored, and as such they play an important role in the selection of targets for desire: an animal's memory of bodily pleasure provides direction for its efforts to eliminate the cause of its pain (35e9-10). But pleasure, unlike pain, does not directly motivate an animal to take action. In this sense at least, bodily pleasure is derivative of bodily pain.

If this interpretation is correct, then Plato's account of bodily pleasure and pain in the *Philebus* is deceptively elegant. He does not deploy EM in the straightforward

fashion that a cursory reading of the textual evidence might suggest. Bodily pleasures and pains are not bodily restorations and destructions; they are *psychological responses* to these restorations and destructions. But EM still applies to bodily pleasures and pains—and so still serves as the centerpiece of Plato's theory—because the relevant psychological responses are *themselves* restorations and destructions. EM thus provides the backdrop for a much richer account of pleasure and pain, one that relies on Plato's extensive claims about the asymmetrical caring relationship between soul and body. In virtue of its caring role, the soul itself has a set of equilibrium states that are sensitive to a corresponding set of equilibrium states in the body; and it is the destruction or restoration of body-sensitive equilibrium states in the soul that constitutes bodily pain or pleasure, respectively.

To my mind, this reading has several virtues. In addition to doing justice to the many twists and turns of Plato's argument, it shows how he can satisfy his express desire to provide an adequate substantive theory of bodily pleasure and pain. In each case the theory specifies a set of essential properties that all and only pleasures (or pains) share, and that serve to explain their characteristic motivational force. Moreover, as I hope to show, this reading allows us to make good sense of Plato's theory of pleasure and pain as a whole. This is an extremely difficult task for any interpreter, since Plato himself never explicitly applies EM—the lynchpin of his theory—to most of the pleasures and pains he discusses. Moreover, as Gosling and Taylor (1982, 132-136) point out, many of these pleasures and pains seem recalcitrant to EM. One important test of my interpretation,

then, is whether it shows how EM can be applied successfully to the other types of pleasure and pain up for consideration in the *Philebus*. My goal in the following sections is to establish that Plato, with the intentionalist position in hand, can apply EM in clear and interesting ways to each of the other types of pleasure and pain he brings up.

1.12. Anticipations

Just after his relatively lengthy discussion of bodily pleasures and pains, Socrates introduces what might be called *anticipatory* pleasures and pains—those that an animal undergoes when it forms a belief to the effect that it *will* be pleased or pained at some point in the future. Socrates claims that these constitute a "different kind" (ἕτερον εἶδος, 32c3-6) of pleasure and pain from the one previously under discussion, which itself also constitutes "a single kind" (ἓν εἶδος, 32b6). It is initially unclear, however, why Plato wants to make this distinction between kinds. Is it his view that an anticipatory pleasure is distinct from a bodily pleasure in part because an anticipatory pleasure is not a restorative process, and hence cannot be analyzed under EM? It is tempting to think so, given the way Socrates introduces the distinction:

> Socrates: Whenever [the harmony in organisms] is destroyed, the destruction is pain; but as for the path to their being, this restoration of them all is pleasure.
> Protarchus: So be it. For that seems to me to be at least a kind of outline.
> Socrates: So shall we endorse as one type of pain and pleasure what happens in each of these experiences [ἐν τούτοις τοῖς πάθεσιν ἑκατέροις]?
> Protarchus: Let it be accepted.

> Socrates: Now about the anticipation by the soul itself of these experiences [αὐτῆς τῆς ψυχῆς κατὰ τὸ τούτων τῶν παθημάτων προσδόκημα], accept also that the prior expectation of pleasures is pleasant and encouraging, while the prior expectation of pains is fearful and painful.
> Protarchus: So this is a different type of pleasure and pain, one that comes to be in the soul itself, in separation from the body, through anticipation?
> Socrates: You've grasped my point exactly (32b2-c6).

In his transition to the second type of pleasure and pain, Socrates seems to suggest that the two types differ, at least in part, because EM applies only to the first. He claims that those of the first type (type-1) consist in the "experiences" of restoration and destruction, whereas those of the second type (type-2) consist in *anticipations of* these experiences. Moreover, as Gosling and Taylor (1982, 135-136) emphasize, Socrates never explicitly entertains the possibility that these anticipations themselves consist in restorations and destructions. So it is tempting to conclude from all this—as Gosling and Taylor do—that Plato intends EM to apply only to the first type of pleasure and pain, and not to the second.

But this conclusion carries a cost. For it means leaving Plato without his best and only candidate for a substantive theory of pleasure and pain—even though, as I argued in the first section, he considers himself obliged to provide just such a theory. As charitable interpreters, we should resist Gosling and Taylor's move, especially if there is a plausible alternative reading of the passage that does not saddle Plato with inconsistent aims.[39] And there is such a reading available. The "experiences" Socrates mentions at 32b7 can be interpreted as restorations and destructions *of a particular sort*, namely, those that the

[39] Cf. Van Riel (2000, 21).

soul (normally) experiences along with the body. On this interpretation, Socrates is saying that the pleasures and pains of the second type are differentiated from those of the first insofar as they are experiences of the soul by itself, and hence do not require—even in normal cases—the body's direct causal influence. This would explain why Socrates and Protarchus repeatedly emphasize, throughout their discussion, that type-2 pleasures and pains are experiences "of the soul itself" (αὐτῆς τῆς ψυχῆς, 32b9; cf. 32c4, 33c5-6, 39d1). Of course they also emphasize that type-2 pleasures and pains are taken in anticipations of type-1 pleasures and pains, and so are second-order in a way that type-1 pleasures and pains are not. But clearly type-2 pains can be both second-order and independent of the body's direct causal influence while at the same time being (psychological) destructions. So it is possible for Plato to distinguish type-2 pains from type-1 pains without denying that they share the property of being destructions. All he needs to say is that they are not destructions *of the same sort*. According to this reading, then, Plato can retain his ambition for a substantive theory without failing to account for the distinction between type-1 and type-2 pleasures and pains. And since this reading is plausible, it is to be preferred over Gosling and Taylor's on grounds of charity.

So far, however, this alternative reading has an edge on Gosling and Taylor's only because it leaves Plato with enough room to hold that type-2 pains are destructions. It gives no additional, text-based reasons to believe that Plato actually does think of them as destructions. But those reasons are available as well. As I have already argued, Plato holds that type-1 pains are psychological destructions that are (in normal cases) linked both

causally and intentionally to bodily destructions. If he wishes to introduce other types of pain without compromising his ambitions for a substantive theory, he can do so—provided that all the tokens of these other types share with all the tokens of type-1 the relevant pain-making property. According to PID, this is the property of *being a destruction of psychological equilibrium*. So if Plato accepts PID, then he is within his rights to claim that the various types of pain differ in virtue of having different causal histories, different intentional objects, or both—so long as he insists that that they do not differ insofar as they are all psychological destructions. Thus his distinction between type-1 pains and type-2 pains does not jeopardize his theory if tokens of the latter type, like tokens of the former, are destructions of psychological equilibrium. And as we shall see, there is an clear sense in which they are.

Shortly after arguing that the source of desire is exclusively psychological, Socrates explores a way in which type-2 pains are like type-1 pains. He begins by asking Protarchus to imagine an animal who is currently suffering a bodily pain, but who "remembers the pleasant things" that would put an end to it (35e9-10). Protarchus suspects that this animal would be suffering a "double" pain—one on account of its current condition, the other on account of its unsatisfied drive to escape from its present condition (36a4-6). Socrates disagrees:

> Socrates: Why do you call this a double pain, Protarchus? Isn't it the case that sometimes someone who is being emptied has a clear anticipation of being filled, but at another time has the opposite—no anticipation at all?
> Protarchus: But of course.

Socrates: And doesn't he seem to you to take pleasure in his memory [of being filled], if he *anticipates* being filled—while at the same time, since he is being emptied, he is in pain?
Protarchus: Necessarily.
Socrates: So that's a time when human beings and the other animals are pained and pleased simultaneously.
Protarchus: It looks that way.
Socrates: But what about when someone who is being emptied does not anticipate getting filled? Isn't *that* when the double experience of pain happens—which you noted a moment ago and thought was double across the board? [ὃ σὺ νυνδὴ κατιδὼν ᾠήθης ἁπλῶς εἶναι διπλοῦν]
Protarchus: That's absolutely right, Socrates (36a7-c2).

In this exchange Socrates claims in effect that an animal suffers a type-2 pain when it expects that one of its bodily desires will not be satisfied. His suggestion seems to be that the satisfaction of an agent's bodily desires is itself the object of some higher-order desire in the agent, such that one's *prospects* of pleasure or pain can themselves be *objects* of pleasure or pain. If this is what Plato wants to get across, then there is a natural sense in which a type-2 pleasure taken in the prospect of a type-1 pleasure is itself akin to a type-1 pleasure: both are psychological states of relief with respect to the well-being of the organism. Type-1 pleasures are states of relief with respect to the *current* condition of the body, while type-2 pleasures are states of relief with respect to the *future* condition of the body. So pleasures can be classified as type-1 or type-2 in virtue of the sort of intentional content they have; but they are all disturbances of psychological equanimity with respect to the body.[40] Moreover, type-1 pleasures are (normally) caused

[40] Tuozzo (1996, 504-508) suggests that type-2 pleasures are not themselves restorations with intentional content, but are contentless feelings *caused by* internal images of restorations. Not only is this interpretation more cumbersome than the one I am developing here; it also attributes to Plato a less philosophically attractive view. Moreover, it is motivated primarily by Tuozzo's antecedent commitment

by the body's current condition, whereas type-2 pleasures are not. So despite their type-specific differences in causal history and intentional content, both of them are restorations of psychological equilibrium with respect to the equilibrium of the body—and this, I take it, makes them amenable to analysis under the intentionalist version of PID. The pleasures and pains of anticipation do not supply counterexamples to the theory I am attributing to Plato.

1.13. Pure Pleasures

In a much later section of the *Philebus* (50e-53c), Plato discusses a type of pleasure that he takes to be importantly different from both type-1 and type-2 pleasures. Socrates variously labels them "unmixed" (50e5-7), "true" (51b1-7), and "pure" (52c1-2) without any apparent shift in meaning. He first introduces them just after distancing his own view of pleasure from the one advanced by some nameless figures he calls "the enemies of Philebus" (44b6). According to Socrates, the enemies of Philebus claim that pleasure itself, in all its forms, is nothing but a release from pain (44c1-2). But Socrates rejects this claim (51a3-4). Though he accepts that the vast majority of bodily pleasures are satisfactions of desire—and hence chronologically presuppose pain-experience of some kind—he recognizes a type of pleasure that one can undergo without suffering any

---

to the idea that Plato has a causal theory of *bodily* pleasure. As I have already argued, however, there are good reasons to deny that he holds such a theory.

previous or simultaneous pain. When Protarchus asks for a more complete account of this third type, Socrates gives the following answer:

> [The pure pleasures are] those related to (τὰς περὶ) the so-called beautiful colors and those related to (καὶ περὶ) shapes, most of those that belong to smells and sounds, and as many of those [pleasures] which, having imperceptible and painless defects, yield perceptible and pleasant restorations (καὶ ὅσα τὰς ἐνδείας ἀναισθήτους ἔχοντα καὶ ἀλύπους τὰς πληρώσεις αἰσθητὰς καὶ ἡδείας παραδίδωσιν) (51b3-7).

In this passage Socrates sets out four examples of pure pleasure and then apparently specifies a property, belonging to each of them, in virtue of which they are all unmixed with pain. His claim seems to be that these pleasures arise from the perceived repair of a defect, where that defect cannot be perceived and so cannot be responsible for any prior pain. (This defect, I take it, is either a result of a previous destruction or an innate state of disequilibrium.) If this reading is correct, then EM obviously applies to the pure pleasures as easily as it does to the impure. In both cases there is a restoration of some sort, and this restoration is tightly linked to the pleasure-experience. The only distinguishing property of a pure pleasure is that the defect it presupposes is imperceptible, and therefore painless.[41]

This interpretation of the passage is contested, however. According to Gabriela Carone (2000, 267n19), the final clause does not specify a property in virtue of which the original four pleasures are supposed to be pure; rather, it specifies an *additional* type of pure pleasure, distinct from the types exemplified by each of the original four. In her

[41] For a similar reading of pure pleasures as painless restorations of equilibrium, see Frede (1993, lv).

view, the four examples of pure pleasure mentioned in the first two clauses are not (linked to) restorations, and so cannot be explained by EM. Carone's interpretation is possible, I think, but strained. As Carone herself recognizes, the structure of the discussion that follows the disputed passage shadows the examples given in it. Socrates moves from shapes (51c1) to colors (51d2) to sounds (51d6) to smells (51e1). At no point does he discuss some further distinct class of pure pleasures—or at least not as such. At 51e7-52b9 he brings up the pleasure of learning, which is the only example he discusses that is not included on the original list. Since there is some lingering doubt about whether these pleasures are really pure (51e7-52a3), he invites Protarchus to apply the criteria set out in the final clause of the passage in order to settle the matter. After some debate, Protarchus agrees that the pleasures of learning are pleasant restorations of imperceptible defects, and hence qualify as pure.

Carone can insist at this point that the criteria to which Socrates appeals here apply only to the pleasures of learning, and not to the other four. But it seems more likely, given the rhythm of the discussion, that these criteria are applied implicitly to all five, and that they are applied explicitly to the pleasures of learning only because there is some question about whether these pleasures belong with the others in a single group. This consideration is not conclusive, however, and if all other things were equal, the contest between Carone's interpretation and mine would rightly be called a draw. But not all other things are equal. Like Gosling and Taylor, Carone leaves Plato with nothing that

could possibly satisfy his self-imposed obligation to provide a substantive theory of pleasure and pain. And this, I think, tips the scales in my favor.

Even if the text justifies the claim that Plato intends EM to apply as much to the pure pleasures as to the impure, it remains something of a mystery how the four examples of pure pleasure that Socrates gives at the beginning of the passage can be understood as restorations. What imbalance in me is being repaired when I smell a rose, listen to a Bach fugue, or gaze at a Mondrian? All of these are obviously perceptual experiences of a certain kind, and so involve or require bodily affections to trigger them. Given that they are caused by "disturbances" in the body—just as impure pleasures are—do they have the same sort of body-oriented intentional content? This seems unlikely.

According to Socrates, the purity of these perceptual pleasures is a function of the transcendent beauty of the objects perceived. He stresses repeatedly that perfect shapes, colors, and sounds are "not beautiful in relation to something [πρός τι], as other things are, but naturally and always beautiful in themselves" (ἀεὶ καλὰ καθ' αὑτὰ πεφυκέναι, 51c6-7); or again, that they are "beautiful, not in relation to something different [πρὸς ἕτερον], but in and of themselves" (ἀλλ' αὐτὰς καθ' αὑτὰς εἶναι, 51d6-9). These formulations are reminiscent of Plato's talk of transcendent forms in earlier dialogues such as the *Phaedo* (74a-75d), the *Symposium* (211b), and the *Phaedrus* (247c-e). But here in the *Philebus* he asserts something he almost certainly would have denied in these earlier dialogues—namely, that one can have a kind of *perceptual* access to such things. If these

are indeed the objects of the perceptual experience that triggers pure pleasure, then it seems unlikely that what is being restored in the perceiver in such cases has any connection at all with the perceiver's bodily equilibrium. For there is no clear sense in which the perception of such entities should affect the soul's attitude toward the well-being of the body it governs. Whatever psychological balance is being restored in these cases, it is presumably not linked intentionally to bodily equilibrium. But it might well be linked intentionally to the equilibrium in other things, such as beautiful sights and sounds. On this alternative view, a pure pleasure is a psychological restoration that is caused by a bodily event, but that takes as its intentional object some instantiation of ideal order. If this is how Plato thinks of the pure pleasures, then they too can be accommodated by PID.

In a later section of the dialogue dealing with the nature of knowledge, Plato suggests—albeit indirectly—that this alternative reading is on the right track. Though he draws several distinctions between different types of knowledge in this section, he also suggests that any type of knowledge worth its name consists in appreciating and applying numbers—and especially proportions (cf. 55e1-3, 56c8-57a2, and 57c9-d2). Later he implies that a thing is *good* just insofar as it has the very proportional order that one can, on his view, come to know.[42] As Socrates puts it, "we should hold [proportional order] responsible for what is in a [proportional] mixture, since [proportional order's] being good is what makes the mixture itself good" (65a3-4; cf. 64d9-e3). Moreover, Plato

[42] Remember that, according to the fourfold division, proportionally ordered items belong to the mixed class and hence are what I have called equilibrium states.

suggests that the mind is something that conforms itself to the character of its objects, especially in the case of knowledge (65d4-10). But if this is what he thinks, then he must also think that as the mind takes on the proportional order of its objects, *it too becomes ordered*.[43] Since even mundane, practical types of knowledge consist in conforming the mind to the order of the world, they improve the souls that possess them.

What makes all this relevant to the question at hand is that Plato takes the pure pleasures to be those that "follow upon [ἑπομένας] knowledge and perception" (66b4-6). To experience the pure pleasure of shape, he thinks, one need only perceive "something straight or round, or what is constructed out of these with a compass [τόρνοις], ruler [κανόσι], or square" (51c4-5). Given that craftsmen employ both "compass" (τόρνῳ, 56c1) and "ruler" (κανόνι, 56b9) in the exercise of their practical knowledge (56b8-c2), they will experience the pure pleasure of shape while they work. After all, it is this sort of instrument that brings *proportional order* to the things being worked up. His suggestion, then, is that the objects of knowledge and perception trigger pure pleasure just insofar as they exhibit proportional order. Moreover, as we have already seen, he thinks that grasping the proportional order in things—via either knowledge or perception—brings proportional order *to the soul*, and hence *improves* it. So there is a clear sense in which the soul of the craftsman, when engaged in his work, undergoes a restorative process. His soul itself is getting ordered: it is *getting into equilibrium*. If this is what Plato thinks, then he can claim that the craftsman's perceptual experience triggers

[43] Cf. Bobonich (1995, 123-136).

in him a pure pleasure—where this pleasure itself is now understood as a restoration of psychological equilibrium. And this is wholly consistent with PID.

1.14. Emotions

In a strange and often overlooked part of the *Philebus*, Socrates takes a great deal of time to discuss what might be called "emotional" pleasures and pains (47e-50d). His primary goal here is to persuade Protarchus that people who are in emotional states such as "anger, fear, longing, lamentation, love, [and] malice" undergo pleasure and pain at the same time. But Socrates never lays out a view about the general nature of emotional pleasures and pains, nor does he elaborate much on the specific emotions he seems most interested in. Because he has so little to say about what these emotional pleasures and pains are, it is initially quite difficult to see how they can be understood as restorations and destructions. My claim, however, is that this difficulty can be overcome.

The lion's share of the section on emotional pleasure and pain is devoted to analyzing the typical state of mind of someone who is watching a comic play. Socrates wants to argue that the audience members, when they laugh, undergo a mixture of pleasure and pain at the same time. He reasons roughly as follows:

(1) Malice (φθόνος) is "a pain of the soul" (λύπην τινὰ ψυχῆς). (48b8-9, cf. 47e1-3)
(2) To be ridiculous is to exhibit the vice of self-ignorance—to believe oneself more rich, more beautiful, more wise, or more virtuous than one really is. (48c4-49c5)

(3) Laughing at something involves taking pleasure in it. (49e9)
(4) The spectators at a comedy laugh at what is ridiculous about its characters. (50a5)
(5) Only the malicious person takes "pleasure in the misfortunes of friends" (ἡδονὴν ἐπὶ τοῖς τῶν φίλων κακοῖς). (50a2-3, 48b11-12)

According to Socrates, this supports his claim that the spectators at a comedy undergo pleasure and pain at the same time:

> The argument says that when we laugh at what is ridiculous about our friends, blending pleasure with malice, we blend pleasure with pain. For we agreed earlier that malice is a pain of the soul, that laughter is a pleasure, and that both of them occur at the same time on these occasions. (50a5-9)

Regardless of its merits as an account of laughter and the ridiculous, this argument indicates a way in which emotional pleasures and pains can be understood as psychological restorations and destructions, respectively. The challenge in this case is to pinpoint a psychological equilibrium that is disturbed by malice and a psychological equilibrium that is restored by laughter. Socrates never explains or clarifies his claim that "malice is a pain of the soul," so it is initially hard to see how this pain is itself a destruction of psychological harmony. But his various other remarks about "the malicious person" seem to suggest that he thinks of malice as a kind of distress at the good standing of one's friends, particularly in comparison with oneself.

There are two main considerations in favor of this reading. First, it is psychologically plausible to suppose that someone who takes pleasure in the misfortunes of his friends is insecure, at some level, about how his own fortunes compare with theirs.

Second, it suits the sense of the term I have translated here—following Frede (1993, 55)—as "malice" (φθόνος).[44] In other dialogues, Plato employs this term to characterize the emotional state of those who would deprive others of various goods (*Phaedrus* 247a7, *Timaeus* 29c2). And he has Socrates claim in the *Republic* (500c1) that malice is nurtured primarily by interpersonal competition. In the pseudo-Platonic *Definitions* (416a), φθόνος is said to be "distress at the goods of one's friends," a view shared by Aristotle (*Rhetoric* 1387b22). All this evidence points in a similar direction: the malicious person is pained by his friends' accomplishments and pleased by their misfortunes because he is insecure about his status as compared with theirs.[45] Since this insecurity can itself be understood as a disruption of psychological equilibrium with respect to one's comparative excellence, malice can be effectively analyzed under PID.[46]

But this presents a complication. How is malicious laughter to be understood as a *pleasure* according to PID? If Plato supposes that taking pleasure in the misfortune of one's friends is a restoration of the very same equilibrium that is disturbed by malice, then his account seems to be in trouble. For in that case he would have to hold that laughing at the misfortune of one's friends is itself the restoration of a sense of security with respect to one's comparative excellence—just as drinking water is a restoration of the body's properly hydrated state. But this is highly implausible. One cannot hope to be healed of one's insecurities by ridiculing others. For this serves only to validate and deepen one's insecurities, not to rid oneself of them. On the other hand, there is

[44] On φθόνος in Plato and Aristotle, see Mills (1985).
[45] Cf. Tuozzo (1996, 510-511).

something undeniable about the claim that malicious pleasure is connected with a kind of relief from some underlying insecurity: if one is insecure about one's own standing with respect to others, one finds something sweet about their failures. Does Plato have a way to make sense of this? Or is he stuck saying that malicious laughter is a restoration of the very same balance that is disturbed by malice? In the following section I will argue that Plato's use of the equilibrium model in his account of comic pleasure and pain is surprisingly sophisticated, and that, as a result, he has the resources to overcome this objection.

1.15. Pains and Mixtures

The only example of emotional experience that Socrates examines closely in the *Philebus* is a spectator's response to comedy. Socrates thinks that this involves a subtle mixture of pleasure and pain, and that such a mixture is common to most (if not all) emotional experience. Indeed, he suggests that his only purpose in bringing up the case of comedy is to show that emotional experience in general has the same "mixed" character:

> Now why do you suppose I showed you the mixture in comedy? Was it not for the sake of persuading [you] that it is easy to display a blend in fears and loves and the rest? And I hoped that once you had grasped this, you would spare me from going on at length about the others and simply take the basic point, that … [they] are [all] full of pleasure blended together with pain. (50c10-d6)

[46] Cf. Frede (1992, 450-451).

So comedy is a relevant topic of investigation here only because it yields a general account of "fears and loves and the rest" according to which they all involve a mixture of pleasure and pain. By "the rest" Socrates means the other five of what I will here call the *seven intense emotions*, which are listed repeatedly throughout the discussion: "we set down anger (ὀργὴν), longing (πόθον), lamentation (θρῆνον), fear (φόβον), love (ἔρωτα), jealousy (ζῆλον), [and] malice (φθόνον) as those in which we claimed we would discover the frequently mentioned mixture [of pleasure and pain]" (50b7-c2; cf. 47e1-3). It seems clear that Socrates wants to establish two things on the strength of the comedy example: first, that each of the seven intense emotions is a mixture of pleasure and pain; and second, that his method for discovering the "mixture in comedy" is suitable for discovering the mixture in all seven cases. But a careful examination of the mixture that Socrates claims to find in the case of comedy shows that his example is not sufficient to accomplish either of these two tasks.

The claim that there is a mixture in comedy depends crucially on the idea that comedies invite us to laugh at, and thereby take pleasure in, the vicious self-ignorance of our friends and neighbors. Because Socrates thinks that we would not enjoy the misfortune of others unless we harbored some profound ill-will toward them, he supposes that this enjoyment constitutes a pleasure of which malice—the last of seven intense emotions listed—must be the cause (τὸν τοῦτ' ἀπεργαζόμενον, 50a2-3). He then concludes, perhaps too quickly, that the pleasure of comedy must be mixed with pain. He claims that he and Protarchus have already agreed that "malice is a pain of the

soul," that "laughing is a pleasure," and that both of them "come to be simultaneously" in the soul of the spectator (50a6-9).

The agreement that malice is a psychological pain actually occurs three times in the preceding discussion (47e3, 48b8, and 49d1). But if Socrates holds to this agreement, then his account has an obvious problem. He wants to show that all seven intense emotions—malice included—are mixtures of pleasure and pain, but his analysis of the so-called "mixture in comedy" reveals that malice is simply a *pain*, not a *mixture*. Indeed, his method seems to show only that malice is a component pain in some *other* mixture. Thus Socrates shows neither that malice itself is a mixture, nor that the method he uses to find a mixture in comedy is suitable for showing that any other of the seven intense emotions is a mixture. So if he is genuinely committed to showing these two things, then he fails.

I suspect, however, that he is not so committed. In the following sections I will argue that, for Socrates, each of the seven intense emotions is not itself a mixture of pleasure and pain, but rather a pain of a very peculiar sort—one that is exacerbated by any direct attempt to relieve it. On this interpretation, the seven intense emotions are pains that generate in us desires whose satisfactions merely deepen the original pains. Though the intense emotions are not themselves mixtures of pleasure and pain, they make such mixtures possible (and indeed likely). What Socrates is getting at here, I think, is that people are driven by emotional pain to do things that only worsen their pain. And in trying to explain such phenomena, I believe he is developing what I will call a *model of*

addiction—a systematic and comprehensive way to make such apparently irrational behavior intelligible.

1.16. Emotional Cause and Effect

One reason to be skeptical of this interpretation is that Socrates, in many of the relevant passages, suggests that he thinks of malice itself as a mixture of pleasure and pain (e.g., 47e1-5; 49c6-d1). If these passages are taken at face value, then it would seem that Plato is unaware of the problem in assuming that the emotions are mixtures, rather than component pains. But these passages are ultimately inconclusive. When Socrates speaks here of "the blend in malice" he may mean nothing more than "the blend associated with malice" or "the blend of which malice is the cause." Moreover, Socrates' treatment of *anger*—another one of the seven intense emotions—suggests that this periphrastic interpretation is justified. When Socrates looks for an example of anger, he turns to Homer's portrayal of Achilles' anger, which Achilles himself claims is both "grievous" and "sweeter than smooth flowing honey" (47e5-48a2). Socrates does not unpack this example, but the context of the passage in the *Iliad* (18.108-109) makes it fairly clear that the grievousness of Achilles' anger is grounded in his perception of having been wronged, and that its sweetness is grounded in his expectation or imagination of getting revenge (cf. Aristotle's *Rhetoric* II.2.1).

If this is how Socrates understands anger, then he thinks of anger and malice in very similar ways. He calls malice "a certain pain of the soul" such that the malicious person "is shown to be pleased at the vices of his neighbors" (ὁ φθονῶν γε ἐπί κακοῖς τοῖς τῶν πέλας ἡδόμενος ἀναφανήσεται, 48b11-12). On this view, malice itself is a pain, but the malicious person is pleased when he apprehends the spectacle of his neighbor's vice. If this is right, then both anger and malice are pains that generate certain specific desires. From the anger of Achilles emerges his desire to see vengeance exacted against Agamemnon; from the malice of the spectator emerges his desire to see his friends and neighbors humiliated. In each case the emotion accounts for a certain desire, but does not itself include the pleasure of *satisfying* that desire. So if there are mixtures of pleasure and pain to be found in the two examples, they cannot simply be the emotions of anger and malice themselves.[47] The mixtures must be *consequent upon* the emotions, since the emotions create the desires whose satisfaction generates the mixed experience. If this interpretation is correct, then it is wrong to accuse Socrates of failing to show that each of the seven intense emotions is a mixture, since this is not something he is trying to do. Rather, he is trying to show that certain mental pains *make possible* a mixture of pleasure and pain, and do so by producing a desire the satisfaction of which is invariably accompanied by pain.

Yet Socrates needs to say more than this. Because he holds that malice itself is an ingredient in the "mixture in comedy," he must hold that malice makes possible a mixture of which it itself is a component. Malice has to be a *component* of the mixture, since

[47] *Pace* Frede (1992, 450).

malice is the mixture's component pain; but it also has to be a *cause* of the mixture, since the spectator's malice serves to explain why he takes pleasure in the spectacle of his friends' and neighbors' vicious behavior. It might seem mysterious how one psychological state could be both a cause and a component of another, and Plato does little to help clarify matters here. But in a previous discussion dealing with mixtures of *bodily* pleasure and pain, he develops the rudiments of what I take to be an interesting general account of such mixtures—whether they be bodily or psychological. As I hope to show, the relevant passages suggest that the primary aim of this general account is to explain how a pain can be the indirect cause of its own deepening.

1.17. Addiction

The section of the *Philebus* on mixtures of bodily pleasure and pain is notoriously cryptic, and to my knowledge no commentator has tried to make sense of it in a sustained and systematic way. This is unfortunate, I think, because if I am right about how the passage should be read, it constitutes a fascinating attempt by Plato to explain the characteristic phenomena of *physical addiction*—and to do so without departing from the framework of EM.[48] Moreover, when this explanation is integrated into his account of the "mixture in comedy," some of the mysteries lingering around that account begin to disappear.

[48] Frede (1993, li) rightly emphasizes that Plato evaluates these pleasures distinctively "from a *medical* point of view." Cf. Frede (1992, 450).

Right from the start Socrates makes it is clear that those who undergo a simultaneous mixture of pleasure and pain are in an unstable and unhealthy condition:

> when someone in restoration or destruction undergoes opposite experiences simultaneously, sometimes he gets hot while shivering and sometimes he gets cold while sweating, seeking, I suppose, to keep one and get rid of the other [ὁπόταν ἐν τῇ καταστάσει τις ἢ τῇ διαφθορᾷ τἀναντία ἅμα πάθη πάσχῃ ποτὲ ῥιγῶν θέρηται καὶ θερμαινόμενος ἐνίοτε ψύχηται. ζητῶν οἶμαι τὸ μὲν ἔχειν, τοῦ δὲ ἀπαλλάτεσθαι]. And this so-called bittersweet mixture, if present and hard to shake, produces irritation and later a savage tension. (46c6-d1)

This passage is clearly intended to describe a kind of feverish condition. Its last sentence implies that one of the two "opposite experiences" here is a pleasure (*qua* "sweet") while the other is a pain (*qua* "bitter"). Now if such pleasures and pains are simultaneous, then according to EM they must somehow be linked to simultaneous restorations and destructions of bodily equilibrium. And since the entire discussion is couched in terms of hot and cold, the restorations and destructions at issue here presumably involve heating and cooling—either of which, of course, might restore or destroy the body's equilibrium temperature. So when Socrates claims that the patient undergoes "opposite experiences" simultaneously, it seems clear that these experiences are "sweet" restoration (pleasure) and "bitter" destruction (pain). This would explain why the patient seeks "to keep one and get rid of the other."

Granted that the patient wants to keep the sweet and remove the bitter, there is still a question about which temperature-changing process is supposed to be sweet, and which is supposed to be bitter. There is an interesting parallel construction in the passage—"he gets hot while shivering" (ῥιγῶν θέρηται) and "he gets cold while

sweating" (θερμαινόμενος ψύχηται)—that mirrors a later construction: "they weep while rejoicing" (χαίροντες κλάωσι, 48a6). From the latter it is clear that the experience to be kept is marked by the participle and the experience to be ditched is marked by the subjunctive.[49] Following this lead, we can read Socrates as saying that the patient seeks to keep "shivering" and stop "getting hot" in the first case, and to keep "sweating" and stop "getting cold" in the second case. So on this interpretation, the "shivering" and "sweating" are restorative processes while "getting cold" and "getting hot" are destructive processes (so interpolate: "getting *too* cold" or "getting *too* hot"). This makes sense intuitively, since shivering and sweating are in fact ways for the body to recover from being too cold or too hot.

One might worry at this point how the same body could be both *getting* too hot and, at the same time, *recovering from being* too hot. The simultaneous restoration and destruction of a single equilibrium state (e.g., proper bodily temperature) in a single body might seem impossible, and there is nothing in this particular passage to show that Socrates is aware of the problem. But in setting out his next example of bodily mixture, he not only seems aware of the problem; he takes steps to solve it. The second example is a case of an itch that does not just go away when scratched. Rather, it intensifies. Like the itch of a mosquito bite, it is caused by an inflammation below the surface of the skin, such that scratching the surface alone only aggravates the itch itself.[50]

[49] Plato's use of μὲν and δὲ here can be understood as calling attention to this.
[50] Dybikowski (1970b, 245) is the only commentator who notices that Plato marks off this sort of itch from the more common sort.

> whenever the seething and the inflammation are on the inside [ἐν τοῖς ἐντὸς], and one does not reach them by rubbing and scratching, but only disperses the surface parts [τὸ δ' ἐπιπολῆς μόνον διαχέῃ], then, by bringing them to fire and to the opposite [of fire] and by alternating one to the other in their distress, sometimes [ἐνίοτε] they produce extraordinary pleasures, but then [τοτὲ δὲ] they produce—by dispersing through force what had been conjoined or in fusing together what had been dissolved—a condition in the internal parts opposite to that of the external parts, such that pains are blended together with pleasures [τοὐναντίον τοῖς ἐντὸς πρὸς τὰ τῶν ἔξω λύπας ἡδοναῖς συγκερασθείσας].[51] (46d7-47a1)

There is an obvious and important distinction at play in this passage between the "surface parts" of the body (or limb) and its "internal parts"—a distinction that turns up three times in the relatively short section on bodily mixtures (46d9-e1, 46e3, and 47c3). That Socrates emphasizes it here should come as no surprise, given that he is trying to explain how the same body can undergo contrary experiences of restoration and destruction at the same time. For it makes sense to say this of a body only if one part of the body is being restored while the other part is being destroyed.[52] So when Socrates claims that an animal can undergo bodily pleasure and pain simultaneously, we should expect him to claim that there are at least two parts of the animal's body, and that the equilibrium in one is being restored while the equilibrium in the other is being destroyed. This is indeed what he claims. The mixture, he says, is one "that comes from blending the surface and internal parts" (47c1-3), and by this he must mean that a restoration in one is blended with a destruction in the other in such a way that the patient's bodily pleasure and pain are simultaneously mixed.

---

[51] For a useful philological discussion of this extremely difficult passage, see Bury (1897, 102-104).
[52] Cf. *Republic* 436b.

But in order to generate this mixture of pleasure and pain, the patient must alternately expose his infected limb "to fire and to the opposite [of fire]." Now by "fire and the opposite," I take it, Socrates means something like "a powerful heating agent and a powerful cooling agent." One might wonder at this point why the patient would first seek to expose his limb to a heating agent, and then change his mind and expose his limb to a cooling agent. Given the model of pleasure and pain that Plato is working with, it seems clear that the patient's motivation to seek either heating or cooling must be based on the perception of the body's disrupted state. If the patient wants to be heated, for example, then he must be feeling too cold somewhere. This suggests that the patient seeks out a heating agent because he supposes (for whatever reason) that his itch will be relieved if he heats it up. Earlier in the dialogue (32a1-8) Socrates claims that the physiological processes of dispersal and thickening are effects of heating and cooling, respectively. Yet he implies in the passage above that the dispersal of the flesh might *also* be an effect of "rubbing and scratching." When one scratches an itch, one disperses (διαχέῃ) flesh that had been overly thickened, thereby bringing about a pleasurable restoration of equilibrium in the surface parts of the infected limb. So a patient looking to disperse his flesh might, on this account, turn to fire: he might believe—falsely, as it turns out—that what he really needs in order to remove the itch is a more powerful dispersal of the flesh than he can achieve by scratching alone.

While this explains why the patient might apply fire to the itch, it does not yet explain why the patient, in a state of "confusion," shifts to the opposite treatment and

then alternates between the two. If it is assumed that the patient's motivation for seeking the "opposite of fire" is similar to the motivation for seeking fire, then the patient switches to the opposite treatment because he now feels, as a whole, overheated. We can explain the patient's new feeling, I think, if we remember that the internal parts of the patient's limb are already unnaturally hot—as Socrates suggests by referring to "the seething and the inflammation on the inside" (46d9-10). Given that the application of fire to the patient's limb winds up heating both its surface and its interior parts, then the mixture of pleasure and pain that the patient experiences as a result is fairly easy to understand. The dispersal produces pleasure on the surface, since it disperses what had been unnaturally congealed; but it produces pain on the inside, since it heats up what was already too hot. So as the surface parts reach the equilibrium state, the inside parts deviate even further from it, until the patient's predominant bodily experience is the pain of overheating. This would lead the patient to shift "in confusion" from fire to the opposite of fire in an attempt to cool the limb down. But because the original "opposition" between surface and subsurface temperature is not removed by either treatment, the second treatment will have essentially the same result as the first: a mixture of pleasure with pain in which pain steadily predominates, until the patient is driven once again to the opposite treatment. Neither of the two treatments cures the disorder, and each feeds parasitically on the failure of the other.

If this is what Plato is getting at, then it is clear how—in his view—a pain might wind up being both a cause and a component of a single mixed experience. The cause of

the itch in this case is the limb's internal "inflammation," which is itself an internal overheating. This inflammation (somehow) causes the surface of the limb to thicken or congeal, thereby bringing the patient to want to "disperse" the surface of the limb. So when the patient applies fire to the limb, he at once restores the equilibrium of the surface and further disrupts the equilibrium of the interior. Thus the pain-component of this mixed experience is simply an exacerbation of the *cause* of the experience itself. By continually intensifying the cause of his anguish, the patient is led to shift wildly from one treatment to the other. And as a result, the patient's condition will worsen dramatically. As Socrates describes it, this shifting treatment will "cause contractions of [the patient's] body, producing leaping and kicking, all sorts of color changes, distortions of features, and palpitations, until it finally drives him totally out of his mind, so that he shouts aloud like a madman" (47a6-9).

What makes this account interesting, rather than just weird, is that it constitutes a fairly sophisticated attempt to explain why certain types of desire-satisfaction merely exacerbate the desire's original cause and so serve only to establish in the desiring person a state of perpetual and all-consuming frustration. If this is the right way to read the text, then Plato is grappling here with the phenomenon of addiction, and—perhaps more importantly—he is trying to give an account of this phenomenon in terms of his equilibrium model of pleasure and pain.[53] Moreover, if Plato thinks that all mixed

[53] Gosling and Taylor (1982, 137-138) acknowledge that one of Plato's goals here is to give an account of desires whose "satisfaction brings about an opposing depletion whose satisfaction in turn reinstates the original desire." But then they claim that Plato's account, "doubtless … is not fully successful. It is not clear, for instance, that thirst, though mixed, is insatiable in the required way." Apparently they fail to see that the sort of experience Plato is concerned with here is not a *diachronic* mixture of pleasure and pain,

experiences exhibit a similar structure, then he presumably thinks that malice is both cause and component of the "mixture in comedy." This would explain why taking pleasure in the humiliation of one's friends and neighbors is not a good treatment for emotional insecurity. For if malice is to the desire for comedy as an internal inflammation is to an itch, then comedy merely exacerbates the malice that makes it pleasant in the first place. Malice leads us to desire our friends' humiliation, but the satisfaction of this desire is accompanied by an intensification of its cause.

This clears away most of the remaining mysteries about Plato's account of emotional pleasure and pain. As I suggested earlier, emotions should be construed as pains in accordance with PID: they are destructions of one's psychological equilibrium with respect to one's own well-being—whether this be understood in terms of interpersonal status (as in the case of malice) or not. And each of them gives rise to a desire the satisfaction of which ultimately deepens the original pain. Plato leaves out the details of how exactly this account is supposed to work, but its general shape is clear enough.

1.18. Virtue

The last type of pleasure Socrates discusses in the dialogue is also the one that seems least amenable to analysis under PID. Yet it receives so little attention from

---

but a *synchronic* one. At no point does Plato suggest that the analysis of addictive pleasures he offers here will apply to standard pleasures, such as those of thirst, which merely presuppose some *prior* pain. His

Socrates that one wonders how much to make of it. This type of pleasure comes up for the first time very early in the dialogue, when Socrates claims that "the temperate person takes pleasure in being temperate itself" (ἥδεσθαι δὲ καὶ τὸν σωφρονοῦντα αὐτῷ τῷ σωφρονεῖν) and that "the intelligent person takes pleasure in being intelligent itself" (ἥδεσθαι δ' αὖ τὸν φρονοῦντα αὐτῷ τῷ σωφρονεῖν, 12d1-4). Since Socrates is interested at this stage only in arguing that such pleasures are intrinsically different from those of the mindless libertine, he does not discuss them any further here. But much later, near the conclusion of the dialogue, he mentions them again. In this section, Socrates stages a kind of mini-dialogue between himself and the various types of knowledge that he and Protarchus have included in the good life. These personified types of knowledge tell Socrates that only some types of pleasure should be included along with them:

> As for those pleasures which you call true and pure, regard them generally as our kin [οἰκείας ἡμῖν]. And in addition to those, mix in [the pleasures] that go with health and temperance, and especially those that always follow upon all virtue everywhere, like attendants upon a god. But as for those that in each case accompany mindlessness and the other vices, it would be irrational to mix them in. (63e3-8)

Socrates deploys the same distinction here that he did at the beginning of the dialogue, but adds that the pleasures of virtue—understood now as forming a class different from that of the pure pleasures—should be included in the good life, while the mindless ones should not.

---

point is that the analysis will work for pleasures that occur with pain *at the same time*.

The challenge here is to explain how these virtue-based pleasures can be understood as conforming to PID. It is tempting to suppose that Socrates, by comparing these pleasures to "attendants" who "follow upon" virtue, is trying to indicate that these pleasures are still what Frede (1992, 440) would call "remedial" in nature. They in some sense *aim at* perfect psychological states—and hence presuppose some sort of deficiency—but are not *themselves* perfect psychological states. The apparent problem with this interpretation is that the agent who takes pleasure in being virtuous does not, *qua* virtuous, seem to be in any way deficient. What remains obscure, then, is the sense in which such pleasures could be restorations of an equilibrium that has been disrupted.

As I mentioned before, Socrates says little here to help clear things up, and it is risky to assume too much about what he means. But I think there is a way for Plato to account for these pleasures without abandoning PID. Following Frede's suggestion, we can argue that for Plato the agent who takes pleasure in being virtuous is indeed defective—though not in the way one might assume. The agent cannot be defective insofar as he is virtuous; but he can be defective insofar as he *fails to appreciate that* he is virtuous. On this view, an agent might exercise temperance and intelligence without registering to himself that this exercise is a perfection of his nature. If this is something Plato would agree to, then the pleasures of virtue can be seen to fit nicely into the framework provided by PID: the virtuous agent's appreciation of his own virtuous state is itself a restoration of that agent's psychological equilibrium with respect to his soul's status. So Plato can apply EM in this case without supposing that the virtuous agent is

defective *qua* virtuous; he need only assume that the agent's soul is in some sense concerned with its own status as virtuous, and so is itself brought into order when it grasps that it has achieved this status.

Once again, this is largely a matter of speculation. Socrates simply does not say enough about these pleasures to give us a very good idea one way or another about their nature. But as I have already argued, Socrates is under a self-imposed obligation to provide an account of the one pleasure-making feature that all pleasures share, and EM is the best (and only) candidate on offer in the entire dialogue. This consideration weighs in favor of reading the virtuous pleasures in something like the way I have suggested.

1.19. Conclusion

If my interpretation is on the mark, then Plato is advancing in the *Philebus* a surprisingly sophisticated general account of pleasure and pain. According to this account, pleasures and pains are not only restorations and destructions of psychological equilibria; they are also bearers of intentional content, where this content reflects their intrinsic responsiveness to the destruction, restoration, and maintenance of equilibrium states. Consequently, pleasures and pains can be classified by type or kind in virtue of the sort of content they have. Perhaps the best way to see this is to line up the different types of pleasure and pain that Plato discusses, and then illustrate how their differences can be captured by analyzing differences in their content.

Let S be any agent who is undergoing a psychological restoration, let T be some entity that is available to S's sense-experience, and let the propositions in angle-brackets express the intentional content of S's psychological restoration. Each of the five types of pleasure can then be given roughly the following analysis:

(1) bodily: <S is undergoing a restoration of bodily equilibrium>

(2) anticipatory: <S will undergo a restoration of equilibrium>

(3) pure: <T is in equilibrium>

(4) malicious: <T is out of psychological equilibrium> (where T is a person other than S)

(5) virtuous: <S is in psychological equilibrium> (where S knows this to be true)[54]

This account can now be generalized, I take it, such that the intentional content of a given pleasure-state is not confined by the range of cases that Plato himself discusses in the body of the dialogue. The general account can be given as follows. For any pleased agent S, S is undergoing some psychological restoration R_S such that, for some proposition | p |,

S's pleasure = R_S <p>

As we shall see, it is this general account of pleasure that tempts Plato into thinking that pleasures, like beliefs, can be false.

[54] Obviously this pleasure will qualify as virtuous (rather than "ridiculous") only if its intentional content is satisfied and only if S knows it to be satisfied. So I have stipulated here that both conditions obtain.

2. False Pleasures

2.1. Introduction

One of Plato's clearest ambitions in the *Philebus* is to show that pleasures and pains can be false in roughly the same way that judgments and beliefs can be false. He has Socrates argue at length—against a reluctant Protarchus—that these two apparently different kinds of mental attitude are subject to the same sort of truth-oriented evaluation (36c3-d4). At the beginning of the argument, Protarchus denies that pleasures and pains can be false (36c8-9), and claims that this by itself is sufficient to distinguish pleasures and pains from judgments and beliefs (36d1-2; 37e12-38a2). But in the end he concedes—under pressure from Socrates—that bad people have false pleasures (40c1-3), and that false pleasures are importantly analogous to false judgments (40d7-e4). My aim in this chapter is to reconstruct and assess Socrates' argument for this striking conclusion. I begin by reviewing his attempt, at the beginning of the argument, to establish that a pleasure can be like a belief in having propositional content. I then turn my attention to the core of the argument, and show that Socrates' crucial achievement is to get Protarchus to agree that any mental attitude is properly called false if its propositional content is false. Since Protarchus ultimately realizes that pleasures—understood now as propositional attitudes—can have false content, he is led to concede that some pleasures

can properly be called false. I conclude by giving some reasons to think that Socrates' argument, if sufficiently distilled and developed, is both interesting and sound.

2.2. Content

In order to establish that pleasures can be false in the same way that beliefs can be false, Plato needs to establish first that pleasures can have propositional content. For it is precisely in virtue of having this sort of content that *beliefs* are truth-apt. At a very early stage of the argument, Socrates indicates that he is well aware of this requirement. He draws an analogy between "judging" (δοξάζειν) and "taking pleasure" (ἥδεσθαι), claiming that both of these mental attitudes have objects: in the first case, "what is judged" (τὸ δοξαζόμενον) and in the second case, "that at which the pleasure-taking is pleased" (τό γε ᾧ τὸ ἡδόμενον ἥδεται, 37a1-9). Socrates clearly wants to emphasize here that at least some pleasures are like beliefs in that they are not free-floating raw feels, but are *oriented toward*, *directed upon*, or *taken in* things. And it is in virtue of this, their *intentionality*, that such pleasures have the content that they do.[55] Throughout the

[55] Penner (1970) is the first to notice that one of Plato's primary aims in this section of the *Philebus* is to show that pleasures can be understood as propositional attitudes. I follow his analysis of the relevant texts quite closely. For a variety of essentially concurring opinions, see Frede (1992, 444-445), Cooper (1999, 158-163), Carone (2000, 275), and Delcomminette (2003, 217-218). Gosling and Taylor (1982, 441) are skeptical, claiming *contra* Penner that "Plato's language nowhere suggests a grasp" of the distinction between attitude and content. Yet their skepticism seems to stem exclusively from their concern that Plato has not noticed a distinction between (i) an agent's being pleased that he will enjoy something and (ii) an agent's enjoying the anticipation that he will enjoy something. But even if Plato does not distinguish (i) and (ii), I fail to see how this affects Penner's claim. For in both (i) and (ii), it seems plausible to suppose that the agent has a pleasure-attitude toward a content that is well-expressed by a proposition in the future tense. Tuozzo (1996, 505) also dissents, claiming that for Plato all pleasures are contentless feelings. But he gives no argument for this view beyond reasserting his interpretation of Plato as holding a causal theory of pleasure and pain. And as I have already argued in chapter 1, this interpretation is unsustainable.

dialogue Socrates marks off the intentional object of a pleasure or a pain by putting it either in the dative without a preposition (cf. 12d1-2; 12d3-4; 37a9), in the dative with ἐπί (cf. 37e5-7; 40d7-10; 49d3-4; 54e5-6), or in the accusative with περί (cf. 51b3-5; 51e1-2; 54d4-7).[56] In each case the construction is apparently designed to mark out whatever it is that the pleasure-taker is pleased *about*.[57]

Since Protarchus never objects to the claim that pleasures and pains can have content in this sense, Socrates is never called upon to provide any further justification for it. Though this is a gap in the argument as it stands, I take it that there are very few philosophers inclined to exploit such a gap. Some might deny that pleasure and pain understood as *sensations* are content-bearers; but few (if any) would dispute that pleasure and pain understood as *attitudes* are content-bearers. It seems uncontroversial that pleasure-attitudes are like belief-attitudes in that one can be pleased that *p* even if it is not the case that *p*. One can be pleased *that the war is ending*, for example, even if the war is only beginning and indeed, even if there is no war. Thus at least some pleasure-ascriptions share with belief-ascriptions the feature of being non-truth-functional. Moreover, the non-truth-functionality of these ascriptions is most plausibly explained by the intentionality of the states ascribed, and by the propositional structure of their

[56] Cf. Penner (1970, 175).
[57] Though Delcomminette (2003, 218-219) agrees with the orthodox view about whether Plato recognizes that pleasures have content, he has a highly unorthodox way of understanding the nature of such content. He specifies "the mere fact of taking pleasure" as a pleasure's "form" and "*what is felt as pleasure*" or "[that] which makes me feel *this* or *that* pleasure" or "the pleasure as it is *felt*" as a pleasure's content. I would agree with Delcomminette if he were simply observing that pleasures are individuated by their contents. But I suspect that he is trying to say more, namely, that this observation by itself provides an adequate account of *what hedonic content is*. On his view, it seems, the content of a pleasure is just that feature by which we individuate it. But this clearly won't do. In seeking a general account of hedonic

intentional content. Though Plato seems merely to assume that some pleasures and pains are intentional in this sense, the assumption, I take it, is relatively harmless. In any case, I shall return to this point.

To say that some pleasures are propositional is not to say that all propositional pleasure-ascriptions can be happily expressed within the "pleased that *p*" idiom. One might take pleasure *in* swimming *as* a form of exercise without being pleased *that* swimming *is* a form of exercise. Yet the content of each of these pleasures carries a logical commitment to swimming's *being* a form of exercise; and it is precisely this commitment that makes the content of each of them truth-apt or, what I will assume amounts to the same thing, propositional.[58] Though Plato never actually argues that pleasures can be propositional without being taken in propositions, he seems to assume almost from the beginning that this is a possibility. Consider the following remark, made by Socrates at the very opening stages of the dialogue:

> we say that the undisciplined person takes pleasure, but also that the sensible person takes pleasure in being sensible itself [ἥδεσθαι δὲ καὶ τὸν σωφρονοῦντα αὐτῷ τῷ σωφρονεῖν]. Likewise, the mindless person takes pleasure, being full of mindless beliefs and anticipations [ἀνοήτων δοξῶν καὶ ἐλπίδων μεστόν]; but the thoughtful person takes pleasure in being thoughtful itself [ἥδεσθαι δ᾽ αὖ καὶ τὸν φρονοῦντα αὐτῷ τῷ φρονεῖν]. (12c8-d4)

---

content, the crucial question to ask is not "what sort of pleasure does S have?" but "what is S pleased about?" Answers to the latter question will supply answers to the former, but not *vice versa*.

[58] Frede (1993, xlv) is right to suggest, contra Gosling and Taylor (1982, 429) that "what is enjoyed" in these quasi-propositional cases is "defined by its description." But here she seems not to notice that the object of enjoyment here can fail to be "an assumed fact or state of affairs" so long as it is something (such as an activity) *taken under a certain description*. All that matters is that there is something *about* the object of enjoyment that is being enjoyed; it is not additionally necessary that what is being enjoyed is some fact or state of affairs. Cf. Carone (2000, 275).

An agent's taking pleasure in his virtuous activity *insofar as it is virtuous* is not equivalent to an agent's being pleased *that* his activity is virtuous.[59] Yet both pleasures seem to have the same truth-evaluable content, namely, that the pleased agent's activity is virtuous. Plato does not call attention to this point in the body of the dialogue, but as we will see it plays an important role in his overall argument.

2.3. Perception and Error

The first significant move that Socrates makes in the argument is to urge that pleasure-attitudes, like belief-attitudes, can be *mistaken*. Since Protarchus has already accepted both that there is an analogy between belief and pleasure, and that this analogy supports a distinction between attitude and content in both cases, Socrates is free to claim that a pleasure-attitude is like a belief-attitude in that it is mistaken when it has the wrong sort of content.

> Socrates: If what is believed is mistaken [ἂν δέ γε ἁμαρτανόμενον τὸ δοξαζόμενον ᾖ], then we must say that the belief making the mistake is not correct [οὐκ ὀρθὴν], and does not believe correctly?
> Protarchus: How could it?
> Socrates: But what if we notice that a pain or pleasure is mistaken about what it is pained or pleased at [ἂν αὖ λύπην ἤ τινα ἡδονὴν περὶ τὸ ἐφ' ᾧ λυπεῖται ἢ τοὐναντίον ἁμαρτάνουσαν], shall we then call it correct or give it any other fine names?
> Protarchus: That would be impossible, if in fact pleasure can be mistaken.

[59] See also Irwin (1995, 319-320), who argues persuasively that in this passage Plato "draws attention to the distinction between the cause and the intentional object of a pleasure. It is true to say both that the cause of an intemperate person's pleasure is an intemperate action and that it is a physical pleasure; but only the second description picks out the intentional object of the pleasure by saying what he enjoys about the action." *Contra* Gosling (1975, 215-216).

> Socrates: Yet it certainly seems that pleasure often arises in us not with correct belief, but with false belief [οὐ μετὰ δόξης ὀρθῆς ἀλλὰ μετὰ ψεύδους].
> Protarchus: Certainly. And in such a case we say that the *belief* is false, Socrates. No one would call the *pleasure itself* false (37e1-9).

To his credit, Protarchus is suspicious of the claim that pleasure-attitudes can be mistaken in the same way that belief-attitudes can. But he nonetheless accepts the claim that *if* pleasures can be mistaken in this way, then both belief-attitudes and pleasure-attitudes can be evaluated on the basis of a prior evaluation of their contents. To sum up, then, Socrates holds (with or without Protarchus' agreement) that the value of an attitude is simply a function of the value of its content. Protarchus is dubious about this last point, but agrees to the weaker point that pleasures can indeed have content.

Having secured agreement on this point, Socrates turns his attention to the possibility of false belief. He begins by inviting Protarchus to consider a case in which an agent (S) has to decide (κρίνειν), in poor visibility conditions, whether or not some faraway object is a man (38c5-7). Socrates claims that S's decision in this case constitutes a belief (δόξα) and also, if uttered by S, an assertion (λόγος, 38e1-4). He goes on to claim that S's soul is "like a kind of book" (βιβλίῳ τινὶ προσεοικέναι, 38e12-13), since S's experience, in effect, "writes words in [S's] soul. And when what it writes is true, true beliefs and true assertions come to be in [S] as a result of this. But when the writer in [S] writes what is false, the result is the opposite of the truth" (39a3-8). Socrates' point here seems to be that S's perceptual experience gives content to S's belief-attitudes such that the truth-value of this content determines whether or not the relevant belief-attitude is

correct. If the content of S's experience ("what the writer writes") is false, then S's belief-attitude toward this content is mistaken. Or so Socrates suggests.

Socrates further suggests that what makes the content of S's experience false is its failure to represent how things are. After hypothesizing that the object which S is straining to see is in fact a man, Socrates claims that if the content of S's perceptual belief is "that it is a man" (ὡς ἔστιν ἄνθροπος) then S "succeeds" (ἐπιτυχῶς, 38d5-7); and if the content of S's belief is "that it is a statue," then S is "misled" (παρενεχθείς, 38d9-10). Evidently Socrates wants to say that S's belief is mistaken just in case its experiential content fails to represent the world as it is. If the content of S's experience is that *p*, then S forms the perceptual belief that *p*. And when it is not the case that *p*, S's perceptual belief is mistaken *because* it is not the case that *p*.[60]

For reasons that are not entirely clear at first, Plato is at pains to emphasize that S also *visualizes* the content of his perceptual beliefs. He has Socrates introduce "another craftsman" who works alongside the writer on filling out the agent's soul-book: a "painter" (ζωγράφον) who "follows the writer, and paints in the soul pictures of [the writer's] words" (τῶν λεγομένων εἰκόνας, 39b6-7). So whenever S forms perceptual beliefs and makes perceptual assertions, S also "sees in himself the images of these beliefs and assertions" (τὰς τῶν δοξασθέντων καὶ λεχθέντων εἰκόνας ἐν αὑτῷ ὁρᾷ, 39b9-c2).[61] Moreover, Socrates thinks, these internal pictures have the same truth-conditional

[60] Cf. *Theaetetus* (189e4-190a8) and *Sophist* (263e3-264b4). It is obvious that in the *Philebus*—unlike in the *Theaetetus* and the *Sophist*—Plato is not particularly worried about what it is for a belief to be false. Instead, for better or worse, he has Socrates provide a rough and ready commonsense analysis.

[61] As Delcomminette (2003, 225, 229) points out, these internal "images" need not be understood as exclusively "visual." One might call up in oneself an "image" of something in a different sense-modality,

content as the beliefs and assertions on which they are based. The picture of a false belief is false and the picture of a true belief is true (39c4-5). This opening phase of the argument, then, can be represented roughly as follows:

Let there be some agent S such that

(0) S believes that *p*.

The first significant premise of the argument, then, is that any believing agent will also be a picturing agent:

(1) If S believes that *p* then S *pictures* that *p*. (39b3-c2)

Call this act of belief-based picturing a *belief-picturing*, and call the content of this belief-picturing a *belief-picture*. The argument continues:

(2) S belief-pictures that *p*. (From 0 and 1.)

(3) S's belief-picture <*p*> is false just in case S's belief that *p* is false. (39c4-5)

(4) S's belief that *p* is false if and only if it is not the case that *p*. (38b6-39a7)

It follows that:

(5) S's belief-picture <*p*> is false if and only if it is not the case that *p*.

---

such as a sound or a smell. But for whatever reason, Socrates uses only optical language throughout the

The first part of Socrates' argument, then, is designed to give an account of what makes a belief—and the internal picturing that accompanies it—false. One might wonder, however, why Plato feels the need to bring talk of "pictures" and "images" into the argument at this stage. As we shall see, Plato seems to suppose that the pleasures of anticipation have a distinctively *visual* sort of content, in that they are taken in images rather than mere thoughts.[62] Since he wants to do justice to the phenomenology of anticipatory pleasure as he sees it, he makes room for a "painter" who illustrates the work of the "writer" by providing internal images of the agent's anticipations. One might deny Plato's phenomenological claim, but the core of his argument does not turn on this point in any case. All that matters for his purposes here is that a pleasure's content, whether intrinsically visual or not, can also be false.

2.4. False Anticipations

The next stage of the argument begins with Socrates' claim that some pleasures "arise with" (μετὰ ... γίγνεσθαι, 37e10-11) or "follow upon" (ἕπεται, 38b9-10) false beliefs. His suggestion seems to be that pleasures, when they "follow upon" beliefs, have

---

argument. So I will ignore this wrinkle in what follows.

[62] See *Timaeus* (71a-72b), where Plato suggests that the appetites are uniquely motivated by images, and so cannot be constrained by discursive reasoning alone. It is important to notice, however, that this does not prevent the appetites from having intentional content and, in virtue of this, having an authoritative role in the agent's deliberations. For the later Plato, the appetites are not brute bearers of force; they have at least some measure of authority—whatever the *source* of this authority may be.

the *same content* as these beliefs.[63] Protarchus accepts the point, but gamely: "in this case we say that the belief is false, Socrates. No one would ever call the *pleasure itself* false" (37e12-38a2). Instead of responding directly to this objection, Socrates tries to tighten the connection between pleasures and the beliefs with which they are linked. For reasons that are not entirely clear at first, he focuses his attention on beliefs about the future:

> Socrates: So in each of us there are propositions (λόγοι) which we call anticipations (ἐλπίδας) …
> Protarchus: Yes.
> Socrates: … plus the pictured images [of these assertions] (τὰ φαντάσματα ἐζωγραφημένα). And one often sees an enormous amount of gold coming to oneself, along with its many consequent pleasures. Moreover one sees oneself in the picture enjoying oneself immensely (καὶ δὴ καὶ ἑνεζωγραφημένον αὐτὸν ἐφ' αὑτῷ χαίροντα σφόδρα καθορᾷ).
> Protarchus: That's right.
> Socrates: Now should we say that the pictures (τὰ γεγραμμένα) good people have in them are for the most part true, since they are loved by the gods? And that the pictures bad people have in them are the opposite? Or should we not say this?
> Protarchus: We certainly should.
> Socrates: So pictured pleasures are no less present in bad people (τοῖς κακοῖς ἡδοναί γε οὐδὲν ἧττον πάρεισιν ἐζωγραφημέναι) but these pleasures are somehow false?[64]
> Protarchus: Yes. (40a6-b8)

---

[63] Cf. Penner (1970, 174). Moordian (1996, 101-102) rightly points out, *contra* Frede (1993, 38n2), that Socrates *never* claims that these pleasures are *identical* to the beliefs with which they arise. Moordian is also right when he says: "it is not clear that the idea that the pleasure has an intentional object or content is the same as that it is identical to or even defined by the anticipatory belief… the sort of content definitive of fear is not simply the anticipation of some event, but the viewing of that event as bad or to be avoided." But then Moordian goes on to suggest, wrongly in my view, that "it is unlikely that Plato treats pleasure as intrinsically representational and hence capable of falsity in much the same manner as belief" (103). Why does Moordian think this? Apparently he supposes that if pleasures have propositional content, then they *are* beliefs: "if one is willing to hold that certain pleasures … have propositional content, then one is accepting that they are judgments" (101). But this claim is utterly unmotivated. There is no obvious reason why one cannot hold both (a) that a certain pleasure-attitude has the same content as a certain belief-attitude, and (b) that these two attitudes are distinct mental states.

[64] According to Thalberg (1962, 69), this particular claim indicates that Plato "confuses the falsity and the wickedness of pleasures." But I fail to see the alleged confusion. Plato's argument turns on an inference from the claim that bad people have false anticipatory pictures to the claim that *the pleasures they picture* are false. Whatever the merits of this inference, it involves no unsupported move from the claim that some pleasures are false to the claim that these pleasures are bad.

This is the first time in the dialogue that Protarchus explicitly agrees to the claim that some pleasures are false. But it is difficult to see how his agreement is justified, given the structure of the argument as Socrates presents it here. For the argument as it stands clearly fails to show that bad people actually undergo pleasures that are false. It trades on an ambiguity between (A) the future pleasures that are anticipated by bad people and (B) the pleasures bad people undergo when they anticipate future pleasures. The difference between (A) and (B) is, in effect, the difference between the pleasures an agent S pictures himself getting and the pleasure S gets by picturing himself getting these pleasures. The problem with the argument as it stands is that it picks out as false only (A), not (B). The only pleasures that Socrates claims are false here are the so-called "pictured pleasures" (ἡδοναί ... ἐζωγραφημέναι). Yet these are merely anticipated by bad people and so are not themselves actual hedonic experiences.[65] Therefore the argument is invalid, since it establishes only that an *anticipation* of pleasure can be false, not that an *actual* pleasure can be false.

One possible way to save the argument as it stands is to assume that it is not intended to establish that actual pleasures can be false. On this interpretation, the argument concludes that bad people have false pleasures in the sense that they will not have some of the pleasures they expect they will have. This falls short of admitting that

[65] Dybikowski (1970a, 152), Penner (1970, 176), and Gosling and Taylor (1982, 437-438) all notice this apparent problem. Gosling (1975, 217) suggests that "painted pleasures" here can simply mean "paintings of pleasures," but his only point in doing so is to indicate that Plato is confused. As I will suggest shortly, there is no good reason to charge Plato with this confusion. Cf. Penner (1970, 176n12).

there are any actually experienced false pleasures, but at least it is warranted by the argument. If this is the right way to understand it, then the argument to this point runs roughly as follows:

(6) Bad people belief-picture that they will get rich and enjoy their riches. (40a3-12)

(7) Bad people are hated by the gods. (39e10-40a1)

(8) The belief-pictures in people who are hated by the gods are generally false. (40b2-4)

(9) Bad people belief-picture that they will get rich and enjoy their riches, and these belief-pictures are generally false. (From 6, 7, and 8.)

(10) Bad people generally will not get rich and enjoy their riches. (From 5 and 9.)

Understood in this way, Plato's argument serves to establish that the pleasures the bad person anticipates getting are false, but only in the sense that *they will not occur*.[66] The obvious problem with this reading is that Plato has not yet shown what he evidently wanted to show, namely, that pleasures can be false in the same way that

[66] They will not occur because the gods make it the case that they will not occur. Plato seems to be assuming here that bad people generally will not get rich and enjoy their riches because the gods will (somehow) prevent this from happening. Likewise, and for similar reasons, good people *will* get rich and enjoy their riches. One might object at this point that Plato would regard a theological view of this sort implausible, since he is sensitive to the fact that wealth and prosperity are not reliable indicators of virtue. The evidence for this is mixed, however. In book X of the *Republic* (612b-614a), Socrates persuades Glaucon and Adeimantus that the just man is loved by the gods, and so will secure—in *this* life—the most of the good things and the least of the bad. Now I admit that this is the conclusion of a pretty lousy argument, and perhaps Plato does not expect us to take it seriously. But even so it shows that Plato is not above exploiting a crude theodicy of this sort when it suits his philosophical purposes. By the same token, we are not obliged to think that Plato is committed to the theodicy on display here. On my reading, the theological hypothesis plays an exclusively *procedural* role in the argument: its goal is to provide a rule for assigning a truth-value to any present anticipation. If so, then it serves merely to specify a class of false anticipations that can serve as hedonic contents. The rule itself—and in particular its theological character—is arbitrary and inessential. Thus my reading saves the evident inferential structure of the argument without loading Plato down with any untoward theological views. For a different but largely congenial account, see Gosling and Taylor (1982, 442-443).

beliefs can be false. If only *pictured* pleasures can be false, then false pleasures are not really pleasures at all: they are merely features of a false picture. This is clearly not enough for Plato, since he needs to provide some account according to which S can be pleased such that *this very pleasure-taking* is mistaken—just as S can form a belief such that *this very believing* is mistaken. An account of this sort is beyond the range of the current argument. For even if there is some intelligible sense in which a pleasure that is not going to happen could itself be mistaken, this is evidently not the sense that Plato needs in order to support his analogy between pleasure and belief.

So if this were the whole of Plato's argument, I think we would be right to regard it as a failure, especially in light of Plato's own broader aims. But I doubt that this is the whole of the argument. Shortly after Protarchus admits that pleasures can be false in the limited sense outlined above, Socrates revisits the analogy between pleasure and belief. His central claim here is that pleasures, like beliefs, can take as their objects states of affairs that do not obtain.

> Socrates: Whoever believes at all is always and in every case really believing, though sometimes not in what is, nor in what was, nor in what will be [μὴ ἐπ' οὖσι δὲ μηδ' ἐπὶ γεγονόσι μηδὲ ἐπ' ἐσομένοις ἐνίοτε].
> Protarchus: Certainly.
> Socrates: And these, I take it, produce false belief and believing falsely. Right?
> Protarchus: Yes.
> Socrates: So should we not give to pains and pleasures the feature that corresponds to these?
> Protarchus: How so?
> Socrates: Whoever is delighted in anything at all in any way is always genuinely delighted, but sometimes not in things that are nor in things that were [ἐπὶ τοῖς οὖσι μηδ' ἐπὶ τοῖς γεγονόσιν] nor, perhaps most often, in things that will be [ἐπὶ τοῖς μέλλουσί].

Protarchus: Necessarily so, Socrates.
Socrates: So wouldn't the same account [ὁ αὐτὸς λόγος] hold, about fear and anger and all things of this sort, that all such things are also sometimes false?
Protarchus: Certainly. (40c8-e5)

According to Socrates, pleasure and belief are analogous in that one can take pleasure in "what is not" just as one can *believe* "what is not." In light of the previous section of argument, it seems clear that Socrates is employing what is sometimes called the *veridical* sense of εἶναι (the verb "to be") in this passage. If so, then by "what is not" he means something like "what is not *the case*" or "what is not *true*." So the upshot of his remarks is that pleasure, pain, fear, and anger are like belief in that they can have false propositional content. What is believed, what is feared, what angers, and what pleases can all be false. Moreover, since it seems unobjectionable to call a belief (derivatively) false when *what is believed* is false, it seems equally unobjectionable to call a *pleasure* (derivatively) false when *what pleases* is false.[67] Likewise for propositional cases of pain, fear, and anger. If this is Plato's point, then it constitutes a crucially important extension of the first part of his argument, for it does something that the first part does not: it explains how an *actually experienced* pleasure can be false. It says that if an actually experienced pleasure is taken in what is not the case, then that pleasure is false.[68] On the evidence of this passage, then, we can build into the argument a further premise:

[67] *Pace* Frede (1992, 445-446), who seems to resist the idea that what is false in this case is the content of a pleasure, rather than the pleasure itself. In her view, "the whole enjoyment provided by the soul's inner dialogue or the painter's work consists in nothing but the logoi or pictures in the soul." But this, again, is to obliterate the distinction that Plato has drawn between pleasure-attitudes and their contents.
[68] In her commentary on the *Philebus*, Frede (1993, 44n2) follows Gosling and Taylor (1982, 435) in suggesting that the purpose of Plato's argument is not to show that a pleasure is false when its

(11) If S is pleased that *p* when it is not the case that *p*, then S's pleasure is false. (40c8-e5)

With this premise in place, Plato is in much better shape. For now he needs only to establish a point that is more or less implied by the earlier part of his argument, namely, that bad people take pleasure in something that is not the case. Given his description of bad people's anticipations, it seems intuitively correct to say that these people take pleasure in what is not the case, since they take pleasure in belief-pictures that are false. And Socrates himself seems to make precisely this point. Just after Protarchus agrees to his claim that bad people's "pictured pleasures" can be false, Socrates infers that "bad people generally enjoy false pleasures" (ψευδέσιν ἄρα ἡδοναῖς τὰ πολλὰ οἱ πονηροὶ χαίρουσι, 40c1). Since the "false pleasures" he refers to here are presumably the "pictured pleasures" that Protarchus has just admitted are false, Socrates seems to be saying that bad people *take pleasure in* these "pictured pleasures"—even though these "pictured pleasures" are features of a false picture. So when bad people belief-picture that they will get rich and enjoy their riches, they also take pleasure in *this*

---

propositional content is false, but rather to show that "the *moral content* of foolish pleasures is mistaken, so that they represent a skewed view of life." Cf. Hackforth (1972, 73): "Plato is doubtless hinting at false-value-judgments, which spring not from the weakness of our bodily eyes but from the blindness of our spiritual vision." Neither Frede nor Hackforth argue for this suggestion, so it is difficult to develop a precise objection to it. But their thought seems to be that bad people's internal pictures of the future are false in part because bad people have the wrong concerns or values. Since getting rich and enjoying one's riches does not *morally merit* being an object of one's pleasure, the pleasure one takes in the anticipation of these events is false. This reading of the argument is important because, as I will argue shortly, it is an accurate interpretation of Plato's ultimate and best-considered view about what makes a pleasure bad. It fails, however, as a reading *of this particular argument*. For the text simply does not support the claim that Socrates is testing the bad person's anticipation for *moral* adequacy rather than, as the analogy between belief and pleasure suggests, (something on the order of) *epistemic* adequacy. As I read it, the argument clearly turns on the question of whether the content of the bad person's anticipatory pleasure is true in the mundane sense. The *moral status* of this content, whether it be true or false, is not Plato's concern here.

very belief-picture.[69] This further claim can now be added to the argument as an additional premise:

(12) Bad people take pleasure in belief-pictures of the type <I will get rich and enjoy my riches>. (40c1)

Having thereby established that anticipatory pleasures are taken in belief-pictures, Socrates can go on to assert that the content of an anticipatory pleasure is nothing other than the belief-picture itself. Though Socrates never explicitly makes this assertion, I take it that the assertion itself is uncontroversial. If S takes pleasure in his belief-picture that he will win the lottery, it would be odd—and indeed, incorrect—to say that S is pleased *that he believes* that he will win the lottery. Better to eliminate the act of belief from the content of the pleasure, and say simply that S is pleased *that* he will win the lottery. Cases of this sort strongly suggest that the pleasure one undergoes in virtue of believing something takes on, by itself, the content of the relevant belief. This is an important move in the argument, because without it Socrates cannot show that bad people have false pleasures insofar as they take pleasure in belief-pictures that are false. What Socrates needs is a further premise to the effect that, for any pleasure H taken in a belief-picture that *p*, H itself, like the belief, has the content *that p*.

(13) If S takes pleasure in a belief-picture <*p*>, then S is pleased that *p*.

[69] This way of reading the argument saves Plato from the charge—leveled by Gosling (1975, 215-219) and Gosling and Taylor (1982, 438)—of conflating not only the enjoyment of picturing with the act of

The rest of the argument, as I understand it, develops naturally with the support of premises 11, 12, and 13:

(14) Bad people are pleased that they will get rich and enjoy their riches. (From 12 and 13.)

(15) Bad people are pleased that they will get rich and enjoy their riches, but generally it is not the case that they will get rich and enjoy their riches. (From 10 and 14).

(16) Bad people have pleasures that are generally false. (From 11 and 15.)

If this reading is correct, then the argument as a whole is designed to establish that a pleasure is false when its propositional content is false. The point of the analogy between pleasure and belief, then, is not that pleasures *just are* beliefs, but that pleasures can take their *propositional content* from beliefs—and can thereby inherit the truth or falsity of these beliefs. This interpretation makes good sense both of the argument itself and of Socrates' own appraisal of it. In his view, the argument has shown Protarchus how "true and false beliefs affect pains and pleasures with their own condition" (42a7-9). And this, I take it, is just to say that the argument shows how (and why) a pleasure is false when it receives its content from a false belief.

---

picturing, but also the act of picturing with the picture itself. On the interpretation I am offering here, Plato is not guilty of either confusion.

2.5. A Brief Assessment

Under this interpretation, Plato's argument is both interesting and defensible. It establishes that pleasure-attitudes, understood as content-bearing mental states, have the same sort of structure as belief-attitudes in that they can bear truth-apt intentional content. Thus they can be false in the same sort of way that beliefs can be false. Moreover, if we leave aside Plato's potentially troublesome distinction between internal pictures and internal assertions, we can reformulate his argument such that it becomes exceedingly hard to dispute:

(a) S can be pleased that *p* when it is not the case that *p*.

(b) If S is pleased that *p* when it is not the case that *p*, then this pleasure is false.

Therefore,

(c) S can be pleased such that this pleasure is false.

This version of the argument seems to be sound. Since Plato arrives at premise (b) by simply stipulating what it is for a pleasure to be false, the only two clear routes of resistance to the argument involve an attack on premise (a). The first route is to deny that any pleasure, as such, can have propositional content; the second route is to hold that one cannot be pleased that *p* when it is not the case that *p*. Neither of these routes, however, is particularly attractive.

To hold that all pleasures are "raw feels" without propositional content is to deny what seems to be obvious, namely, that one can be pleased *about* something. It seems quite possible to be pleased *that* Nixon is dead, *that* the Cowboys won, or *that* the Pittsburgh Symphony will be coming to New York. A skeptic might dispute the status of such evidence by claiming that the sort of talk involved here is systematically misleading, and that each of the above expressions is shorthand for a more complex and accurate expression in which reference is made both to a belief and to that belief's hedonic effects. But this way of dealing with the issue is a nonstarter. It is one thing to claim that the best account of hedonic content is a causal one; it is quite another to claim that pleasures cannot have content at all. In order to dispute premise (a), the skeptic cannot rest easily with the first claim, since (a) can be true even if a causal account of hedonic content is the best one available. The skeptic has to claim instead that we cannot be pleased *about* anything, even if a thoroughly causal account of this "aboutness" were in the offing. And this claim, as I mentioned earlier, involves what seems to be a denial of the obvious.

The other route of resistance to (a) is to hold that, necessarily, if one is pleased that *p* then it is the case that *p*. This seems a close approximation to the idea that drives Protarchus, at least initially, to oppose the argument. Right from the start of the discussion, he conspicuously fails to dispute Socrates' claim that pleasures are truth-apt, apparently because he thinks that all pleasures are *true*.[70] His reasons for thinking this, however, are at best unclear. Once one admits that pleasures can be propositional, it is hard to see what grounds one might have for denying that the content of such pleasures

could be false. Perhaps Protarchus assumes that pleasure-taking is like perception, in that the hedonic content, like perceptual content, is never false. If an agent perceives that *p*, Protarchus might say, then it is the case that *p*.[71] But whatever the merits of drawing an analogy between perception and pleasure, this particular use of the analogy is unhelpful. For there is no principled reason to suppose that taking pleasure is like perceiving *in that* taking pleasure involves a successful grasp of something's being the case. Moreover, since the same perceptual content can figure in both a perception and a *mis*perception, it seems wrong even to say that perceptual content, in this sense, is never false.

So premise (a) of Plato's argument is strong enough to withstand the two primary lines of resistance to it. Is there any promising line of resistance to premise (b)? As I mentioned before, the problem with disputing this second premise is that it seems to be true by definition. According to Plato, what it is for a pleasure to be false (in the sense canvassed by this argument) is for it to be the bearer of false propositional content. Yet the skeptic need not be cowed by this into thinking that Plato's definition cannot be challenged. The stipulative status of (b) might ensure that the argument as it is presented is sound, but the skeptic might claim that the argument's conclusion establishes something far weaker than what Plato really wants. Perhaps the best way to bring out this line of criticism is to recall the distinction, trotted out by Socrates at an early stage of the argument, between a mental attitude and its content. There Socrates makes it clear that the overall goal of the argument is to show that pleasure-attitudes are like belief-attitudes

[70] Cf. Moordian (1996, 94).

in that they can be mistaken just in virtue of having the wrong sort of content. Now the skeptic might agree, for the sake of discussion, that a belief-attitude is mistaken just in virtue of having false content. But then he might insist that, for all he can tell, Plato has not shown that a *pleasure*-attitude is mistaken just in virtue of having false content. All he has shown is that pleasure-attitudes can have false content; he has not shown that a pleasure-attitude's having false content is a way for it to be *in error*—or at any rate, for such an error to be a bad thing *per se*. To get what he wants, and to get around the skeptic's objection, Plato needs to supply (and defend) a further premise:

(E) If a pleasure is false, then it is mistaken.

As we shall see in what follows, Plato gives us no good reason to think that (E) is true—at least not in this section of the *Philebus*.

2.6. Evaluating False Pleasures

In four of his most celebrated dialogues—the *Protagoras*, the *Gorgias*, the *Phaedo*, and the *Republic*—Plato advances lengthy arguments to the effect that our pleasures and pains are appropriate targets of ethical criticism.[72] But in the *Philebus* he develops this thought in a striking new direction. He suggests, somewhat brazenly, that

[71] This roughly follows Moordian's (1996, 105-112) speculations about what Protarchus' grounds might have been.

our pleasures and pains should be criticized in precisely the same way that our beliefs and assertions are. Indeed, he has Socrates claim that the only way to explain why a particular pleasure is *bad* is to show that it is *false*. I will call this view the *Grounding Thesis*, since it holds that a pleasure's falsehood is always the ground for its badness.[73]

Socrates first gives voice to the Grounding Thesis just after he has persuaded his hedonist friend Protarchus not only that it is intelligible to speak of a pleasure's being false, but that most people suffer false pleasures on a regular basis. The argument Socrates uses here hinges on a tight analogy between pleasure and belief, and it is to this analogy that Socrates returns when the argument is complete. His intention, apparently, is to extend the analogy as far as it can possibly go:

> Socrates: Well then, do we have any way of telling bad beliefs from good ones other than by their being false?
> Protarchus: No.
> Socrates: Nor, I think, do we grasp [κατανοοῦμεν] any way for pleasures to be bad except by *their* being false.
> Protarchus: But you've got it completely wrong, Socrates! It's not in virtue of their falsehood that pains and pleasures are thought to be bad, but in virtue of some other great and wide-ranging badness they fall in with [συμπιπτούσας]. (40e6-41a4)

According to Socrates, then, pleasures are like beliefs in that their being bad can be explained only by their being false. This view can be usefully expressed as follows:

---

[72] See *Protagoras* 351b-358a; *Gorgias* 492d-500a; *Phaedo* 64d-67b; and *Republic* 580d-588a.
[73] Of the many commentators who grapple with the problem of false pleasure in the *Philebus*, only Dybikowski (1970a, 160-161) takes explicit notice of Socrates' espousal of the Grounding Thesis. He suggests that "the development of some such thesis would appear to be the final object of an insistence upon the legitimacy of speaking of pleasures as false" (161). But he does not attempt to lay out how the thesis could be developed, nor does he venture any evaluation of it as it stands.

(GT) For any pleasure H,

(G1) if H is bad, then H is false, and
(G2) if H is false, then H is bad *because* H is false.

Protarchus, to his credit, suspects that GT overextends the analogy between pleasure and belief. Though he is willing to concede that pleasures and pains can be false (40e2-5), he chafes at the further claim that a pleasure or pain is bad just in case—and *because*—it is false. In his view, what makes a pleasure or pain bad is not its being false, but its having some very different sort of property that has not yet been discussed.

Instead of responding directly to this worry, Socrates effectively ignores it.[74] He tables the issue and proposes that they consider a new and different way in which pleasures and pains can be false (41a5-b2). This is a shame, because Protarchus' objection is challenging and highly intuitive. After all, GT does seem implausible, and for roughly the same reason that Protarchus gives: if and when we submit a person's pleasures and pains to ethical scrutiny, we do not seem to base our assessment of them on what we take to be their *truth*-value. To respond adequately to this objection, Plato needs to show more clearly how his account of false pleasure underwrites his claim that falsehood is the only bad-making feature that pleasures can have.

My aim in the rest of this chapter is to evaluate Plato's prospects for accomplishing this task. I begin by arguing that the account of false pleasure Plato develops in his argument from false anticipations cannot, by itself, support an adequate defense of GT. According to this account, remember, a pleasure is false just in case its

intentional content is false. In my view, this account fails to vindicate GT in two connected ways: first, it does not adequately reflect the ethical standards that pleasures are expected to meet; and second, it muddles the distinction between pleasure and belief. After criticizing Plato's main account of false pleasure on these grounds, I turn my attention to a later and often-overlooked part of the *Philebus* (54e1-55a11), where Plato unveils the core of what I take to be a different and much-improved view. On this new view, roughly, a pleasure is false just in case its propositional content fails to represent a good state of affairs—that is, a state of affairs the obtaining of which would be good in itself. I argue that a suitably developed version of this second account can remedy the deficiencies of the first one and, as a result, yield a coherent and interesting way for Plato to parry Protarchus' objection and vindicate GT.

2.7. Falsehood and Error

If Plato is to defend GT successfully, it is not enough for him to show that pleasures can be false in some sense; he also needs to show that a pleasure's being false in that sense is the only ground for its being bad. This may be the reason why he has Socrates persuade Protarchus, at a very early stage of their discussion, that the same formal criterion of badness applies to pleasures, pains, and beliefs alike:

[74] Again, only Dybikowski (1970a, 161) takes note of this.

Socrates: If a mistake is made about what is believed [ἂν δέ γε ἁμαρτανόμενον τὸ δοξαζόμενον ᾖ], then we must say that the mistaken belief is not correct [οὐκ ὀρθὴν], and does not believe correctly?
Protarchus: How could it?
Socrates: But what if we notice that a pain or pleasure is mistaken about what it is pained or pleased at [ἂν αὖ λύπην ἤ τινα ἡδονὴν περὶ τὸ ἐφ' ᾧ λυπεῖται ἢ τοὐναντίον ἁμαρτάνουσαν], shall we then call it correct or give it any other fine names?
Protarchus: That would be impossible, if in fact pleasure can be mistaken. (37e1-9)

It is important to appreciate what Socrates says and does not say in this passage. He says that pleasures are like beliefs in that they are bad if they are mistaken; he does *not* say that pleasures are like beliefs in that they are bad if they are *false*. If he had made this latter claim, he would have begged the question in advance against Protarchus. By limiting himself to the claim that *error* is the criterion of badness for both pleasure and belief, Socrates leaves himself room to draw a distinction between error and falsehood and then argue (without begging the question) that pleasures, when false, are mistaken—and hence bad.

This distinction between error and falsehood is crucial not only for the purpose of defending GT in a non-vacuous way, but also—as we have seen—for the purpose of understanding propositional attitudes. One cannot hope to understand what it is to believe something, for example, unless one is prepared to distinguish between a belief and its content. *What is believed* in a given case—that is, the content of the belief—might also be what is wished, what is imagined, what is pleasing, and so on. (One can believe, wish, imagine, or be pleased *that it is raining*, for example.) Belief is just one of many possible

attitudes a person can take toward a specific content. What seems to make the attitude of belief special is that it involves a commitment to the *truth* of that content. This is why it makes sense to say that a belief is mistaken, and hence bad, just in case its content is false. To distinguish between error and falsehood in this way is to respect the distinction between attitude and content, and to affirm that the two must be evaluated differently (though in the case of belief, of course, these evaluations are tightly linked.)[75] So what Plato needs to show, given this distinction, is that pleasure-attitudes are vulnerable to the same sort of truth-oriented norms that belief-attitudes are. And he can do this only if he can show that pleasure-attitudes, like belief-attitudes, carry an exclusive commitment to the truth of some content or other. For only then is he entitled to claim that a pleasure is mistaken (and hence bad) just in case, and *because*, it is false.

2.8. Some Problems for the First Account

The challenge for Plato, then, is to give an account of false pleasure that picks out, for each pleasure, a content the truth of which that pleasure is plausibly committed to. Yet it is highly questionable whether the first account of false pleasure on offer in the *Philebus* can meet this challenge. On the first account, remember, a pleasure is false just in case its intentional content is false. More formally, then,

[75] My distinction between attitude and content is effectively equivalent to Penner's distinction between "process" and "product" (1970, 169). As he puts it, belief "may mean something I do or tend to do, namely believing, or it may mean something believed by me, just as cooking may be something I do (a process) or something cooked by me (a product).… 'True' and 'false' apply primarily to the latter, but are applied derivatively to the former." Cf. Gosling and Taylor (1982, 437).

(F1) For any pleasure H with content <*p*>, H is false just in case <*p*> is false.[76]

If Plato enlists F1 as an analysis of false pleasure in his attempt to defend GT, then he has to defend the following two claims:

(F1/GT) For any pleasure H,

> (F1/G1) if H is bad then H's content is false, and
> (F1/G2) if H's content is false then H is bad because H's content is false.

The main problem with F1/GT is that it is highly implausible. Consider F1/G1, its first component. F1/G1 seems wrong because not all of the pleasures that Plato (and we) would classify as bad have false content. Consider, for example, intense bodily pleasures such as those associated with vigorous indulgence in eating, drinking, or sex—pleasures which, according to Plato, should be avoided at all costs (45a4-e7, 63c5-64a6). On the assumption that these pleasures are mere sensations, rather than propositional attitudes, they are not apt for being false on F1. But then they are bad, according to Plato, without being false. So if Plato thinks that bodily pleasures are non-propositional, then he cannot sustain F1/G1.

Yet even if he denies that such pleasures are non-propositional—as I argue he does in chapter 1—he runs into another, more serious problem. In his view, remember, bodily pleasures (normally) take restorative bodily processes as their intentional objects, and so are as logically liable to being false on F1 as any other propositional pleasures are.

Roughly speaking, their content takes the following form: <a bodily equilibrium is being restored>. Yet even on the assumption that all intense bodily pleasures are both propositional (in this sense) and bad, it is evident that not all of them have false propositional content. Bodily pleasures with false content are those that register a restoration where, in fact, no restoration is taking place. But then the vast majority of intense bodily pleasures do not have false content. So if no bodily pleasure can be bad unless it has false content—as F1/G1 holds—then almost all intense bodily pleasures are not bad. And this contradicts Plato's own view.

A similar problem for F1/G1 arises in the case of non-bodily propositional pleasures. Consider the psychological phenomenon of *Schadenfreude*, where some agent S is pleased that another agent T is suffering some misfortune or other. Plato has Socrates condemn Schadenfreude as "unjust" in most cases (49c8-d10), and rightly so. Ethical intuition strongly suggests that there is something intrinsically bad about S's pleasure here, even (and especially) if T really is suffering the relevant misfortune. But if this is right, then it seems obvious that S's pleasure is bad whether or not T is suffering, and hence whether or not S's pleasure has false content. The ethical status of Schadenfreude is utterly botched by F1/G1, since F1/G1 holds that S's pleasure is bad just in case T is not actually suffering—which is absurd, even by Plato's own lights.

[76] I use the somewhat clunky phrase "with content <*p*>" here because, as we have seen, not all pleasures with propositional content are well-expressed within the "pleased that *p*" idiom.

2.9. The Grounding Thesis Modified

Given these difficulties, Plato must either abandon his main account of false pleasure (F1) or modify GT by dropping the claim that pleasures are bad only if they are false (G1). To make this move is not so much to reject the Grounding Thesis as to weaken it. For Plato can still claim that if a pleasure is false, then it is bad because it is false (G2). And G2, if true, is enough to establish the desired explanatory link between a pleasure's falsehood and its badness—without implying that every bad pleasure is false. If Plato exploits this option and retains F1, then he need not defend the hopeless F1/G1. He need only defend F1/G2, the claim that if a pleasure's content is false, then that pleasure is bad because its content is false. This weakened version of the Grounding Thesis, unlike the original GT, has at least a chance of being true under F1.

At first blush, F1/G2 seems to stand on better ground than F1/G1. There does seem to be something bad about experiencing pleasures that have false propositional content. Attempting to explain why this is so, Irwin (1995, 330) suggests that an agent who has such a pleasure is *ipso facto* defective: "In order to have false anticipatory pleasures, [for example,] we must suffer from some recognizable defect, in this case the cognitive defect that gives us the false anticipations. We are justified in preferring to be free of this defect." Irwin is referring here to a section of the *Philebus* (36c-41a) that we examined earlier, in which Socrates asks Protarchus to consider the case of an agent (S)

who anticipates that he will get rich and enjoy his riches (40a9-12). As I argued before, Socrates claims that S's anticipatory pleasure is false just in case either of the following two propositions is false:

(1) S will get rich.

(2) S will enjoy his riches.

If either (1) or (2) is false, so is the anticipation; and if the anticipation is false, so is the pleasure taken in it. As applied to this case, F1/G2 holds that if either (1) or (2) is false, then S's anticipatory pleasure is bad *because* it is false. Since Socrates claims in the course of this argument that "bad people as a rule enjoy false pleasures" (40c1), perhaps his point is that if S's pleasure reflects or expresses a certain defect in S—one that is revealed by S's false anticipation—then S's pleasure is itself defective, and hence bad. Or so Irwin suggests.

One objection to F1/G2, as Irwin seems to understand it, is that it's too rigid. Assume for the sake of argument that S's anticipation is false because (1) is false. Does the fact that S will not get rich constitute good evidence that S suffers from some sort of defect, cognitive or otherwise? It seems not. S might have good reason to anticipate getting rich even if—through no cognitive or other fault of his own—contingent circumstances prevent this from happening. So we cannot safely infer from the falsehood of (1) that S's anticipatory pleasure exhibits a cognitive defect in him.[77] This blunts the

[77] As Gallop (1960, 332) puts the point, "one may take pleasure in a piece of false information when one is in no way to blame for being misinformed."

intuition enlisted in support of F1/G2, since it no longer seems right to say in every case that if a pleasure is false, then it is bad because it is false.

Anthony Kenny (1960, 51-52) notices this problem and tries to amend Plato's view to account for it. He agrees that the falsehood of (1) does not necessarily indicate a deficiency in S, but denies that Plato is particularly interested in cases where (1) is false. In his view, Plato is interested primarily in cases where only (2) is false—where S is mistaken not about what he will get, but about whether he will enjoy it once he gets it (or *would* enjoy it *should* he get it).[78] Kenny thinks that this error is unlike the first in that it involves a failure of self-knowledge rather than world-knowledge. If only (2) is false, then S is not mistaken about how the world will be; he is mistaken only about what he himself enjoys. According to Kenny, this constitutes a cognitive failure with obvious ethical significance, since it implies that S is ignorant of his own real preferences.[79] So Kenny proposes, in effect, that F1/G2 be applied only to pleasures that have appropriately self-oriented content. In such cases, he thinks, the claim rings true.

But Kenny's modification may not help much. To begin with, the distinction between self-knowledge and world-knowledge is not obviously effective in silencing worries about whether S necessarily suffers from a cognitive defect just in virtue of having false anticipations. Just as (1) might be false through no fault of S's own, (2) might be false through no fault of S's own. For it seems entirely possible that S, when he forms his anticipation, has good reason to think that he *will* enjoy his riches. Should something

[78] Kenny draws here on a distinction first explored by Williams (1959, 69).
[79] Cf. Gosling and Taylor (1982, 442).

happen to S in the meantime such that his character changes in ways that make it impossible for him to enjoy being rich, then (2) is false. Yet S's inability to foresee these changes does not obviously reveal something cognitively defective about him. Kenny might reply that it is only the instability of S's character that makes the truth of (2) contingent on unforeseeable events. If this instability is something for which S is responsible, then the falsehood of (2) indicates that S has a defective character which is reflected or expressed by S's pleasure. But to make this move is to preserve the coherence of the view by making it less attractive. For it is not at all clear that stability is more valuable than flexibility in the face of contingent and unforeseeable events.

Whether stability of character is a virtue or not, both Kenny's and Irwin's defenses of F1/G2 suffer from a deeper problem. They both fail to draw a sufficiently robust distinction between the value of S's character and the value of S's pleasure, fueling what I take to be an illicit slide from the claim that S is bad to the claim that S's pleasures are bad. For even if an agent who experiences false self-oriented pleasures is somehow bad, and even if these pleasures reflect the badness of that agent's character, it does not follow that *these pleasures themselves* are bad.[80] Imagine an agent who takes pleasure in the anticipation that he will do something courageous at the Assembly and enjoy it. Now assume that the agent does the courageous thing, but does not enjoy it. According to both Irwin and Kenny, there is something bad about the agent—his self-ignorance, let us

[80] Dybikowski (1970a, 160) sees this: "Socrates makes the point that good men have true anticipatory pleasures, while bad men often experience false ones. But this remark of itself does not show that the pleasures of bad men are themselves bad and the pleasures of good men good, but only that they differ from each other." *Pace* Frede (1985, 175-176).

say—and therefore there is something bad about the agent's anticipatory pleasure. But I fail to see how the pleasure itself is implicated in the disvalue here. If (by hypothesis) the agent takes pleasure in the prospect of enjoying a courageous act, then it does not seem relevant to the evaluation of this pleasure whether the agent really will enjoy being courageous when the time comes. The source of disvalue here seems to be the agent's failure to live up to his self-ideal, not the content of his self-ideal, and not the agent's pleasure-attitude that bears this content. Indeed, the agent's pleasure seems to have redeeming value insofar as it is directed at an ethically appropriate self-ideal. It is a good thing, at least, that the agent takes pleasure in the anticipation that he will enjoy being courageous. Whether or not he actually will enjoy being courageous might be relevant to the evaluation of his character, but it seems irrelevant to the evaluation of his pleasure.

One might object at this point that pleasures taken in doing the right thing, if consistently false, may well play a systematic causal role in *making* themselves false. It seems at least possible for an agent to have anticipations that prevent their own realization precisely because they are objects of pleasure for him. Someone might be so easily excited and distracted by his anticipation of doing the courageous thing that he invariably fails to do it. This case is structurally similar to that of Davidson's (1980, 79) mountaineer, who desires to get out of danger and suddenly comes to believe that he can do so by letting go of his friend's rope. This belief so unnerves him that it causes him to let go of the rope. In a parallel case of this sort, an agent's false pleasures might have systematically bad consequences, and this might be a reason to suppose that they are bad

insofar as they are false. But the moral of this case cannot easily be generalized, since some false self-oriented pleasures presumably are not causally responsible for their being false. Indeed, it is not difficult to conjure up scenarios in which an agent's self-oriented anticipatory pleasures, despite being systematically false, are the only salient mental states that weigh causally in favor of making them(selves) true.[81] So even if some pleasures are causally responsible for making themselves false, there is no good reason to conclude from this that every false pleasure is bad because it causes itself to be false.

Even if it were granted as a matter of causal law that the bad person's anticipatory pleasures systematically falsify themselves, I doubt this is sufficient to show that such pleasures are bad. The thought lurking behind this doubt is that if pleasures are to be evaluated *as attitudes*, then their relations to other states of mind or character are not relevant to their evaluation. The value of a pleasure-attitude, like that of a belief-attitude, hinges exclusively on the attitude that it is and the content that it has. All that matters is whether it is appropriate to have *this* attitude with *this* content. When evaluating a belief-attitude as an attitude, one should consider only whether its content is true. Whatever relations this belief may have to other psychological events or states is strictly immaterial. Indeed, this is precisely why it seems so jarring to criticize an agent's beliefs on grounds that do not bear on the truth-value of its content (e.g., because it will have harmful effects, or because it is the product of a bad character). My claim, then, is that pleasures—if understood as attitudes—are entitled to the same evaluative treatment that

[81] Imagine a student who is terminally lazy, but who every day takes pleasure in the anticipation that he will get his act together and go to class the following day. This pleasure is the only thing that ever rouses

is accorded to beliefs. Facts about the agent's character, and about the way his character is expressed or affected by his pleasures, simply do not bear on the question of their value as attitudes.

2.10. A Different Defense

Instead of arguing (with Irwin and Kenny) that a pleasure with false content is bad because it is the expression of a defective character, one might argue that a pleasure with false content is bad because it, like a belief, carries an intrinsic commitment to the truth of its content. An interesting case can be made for this claim. After all, propositional pleasures often seem to be as sensitive to evidence as beliefs are. If you both believe and are pleased that it is raining, for example, then if and when you are faced with evidence that it is not raining, your pleasure is liable to vanish along with your belief.[82] What this suggests is that a pleasure with content <*p*> is committed to the truth of <*p*> and so is mistaken if <*p*> is false. If this is right, then every propositional pleasure is governed by the same truth-oriented norms that govern belief.[83]

Yet it is hazardous to generalize here, because not all propositional pleasures seem to have this feature. Consider, for example, the pleasures of fantasy. These have propositional content—since they are taken in imagined states of affairs—and normally

---

him to make preparations for going to class, yet these preparations are invariably in vain.

[82] This phenomenon is noted and discussed by Williams (1959, 66), Penelhum (1964, 84), and Penner (1970, 167).

[83] Penner (1970, 167-170) recommends attributing to Plato a view of this sort.

their content is false.[84] But clearly they do not exhibit the same sensitivity to evidence that the standard propositional pleasures do. The pleasure I take in the fantastic thought that I will one day be a US senator is completely insensitive to the overwhelming evidence that I will not. In cases such as this, the pleasure-taker is apparently not committed to the truth of his pleasure's content. And this is just to say that one can be pleased with respect to propositional content <*p*> even if one does not believe that *p*, and indeed, even if one believes that *not-p*. But if pleasure and belief can come apart in this way, then it seems wrong to claim that any pleasure with false content is mistaken. For the agent who merely entertains a pleasing proposition is not committed to the truth of that proposition—and so is not, *qua* pleasure-taker, vulnerable to the norms of belief. Indeed, this is precisely why it seems unreasonable to insist that there is something inherently wrong with the pleasures of counterfactual imagination as such.

In response to this objection, the defender of F1/G2 might try to modify F1 so as not to classify the pleasures of fantasy as false.[85] The most straightforward way to do this is to modify F1 such that an agent's pleasure registers as false only if its content is also the content of one of the agent's beliefs.

(F1a) For any pleasure H with content <*p*>, H is false just in case:

---

[84] Weirdly, Thalberg (1962, 70) denies this: "*Pleasure in the thought of* [where there is no judgment involved] is not a counterexample to Plato's thesis, since mere *thoughts*, or unasserted propositions, are neither true nor false." If Thalberg really means what he says here, then apparently he holds that the proposition <unicorns rule the universe> is not false unless someone asserts it. If, on the other hand, he means that only *judgments* or *assertions* are truth-apt, rather than their contents, then he is—despite himself—contradicting Plato's thesis. For whatever pleasures are, I take it, they are neither judgments nor assertions.

[85] This strategy is pursued most vigorously by Frede (1985, 173-175) and (1993, xlvii-xlviii). See also Thalberg (1962, 71) and Delcomminette (2003, 220-229).

For some agent S,

(i) S has H,
(ii) <*p*> is false, and
(iii) S believes that *p*.

According to F1a, the pleasures of clear-eyed counterfactual imagination are not false, since an agent's pleasure is false only if it shares its content with one of that agent's beliefs. Thus any pleasure that is liable to falsehood on F1a is as sensitive to evidence as a belief is, and so can be construed as carrying a commitment to the truth of its content.

But F1a merely aggravates the problem it is intended to resolve. To claim that an agent's pleasure is false only if its content is also believed by that agent is merely to invite the same Protarchus-style skeptical objection in a different and more developed form. For now it seems that the sense in which pleasures can be false is crucially *not* analogous to the sense in which beliefs can be false. The analogy, remember, was supposed to underwrite Plato's view that pleasures, like beliefs, have a representational function that makes them answerable to truth-oriented norms. But now the defender of F1a/G2 is forced to concede that these norms apply only to a certain subset of pleasures, namely, those that are parasitic on beliefs. The skeptic can exploit this concession by claiming that pleasures carry no truth-oriented commitments *in themselves*, since they are answerable to truth-oriented norms only when they depend for their survival on the beliefs that give them their content. Without beliefs to give them content, pleasures cannot be false. But this is just to say that pleasures carry no intrinsic and belief-independent commitment to truth, and so are not, in themselves, answerable to truth-

oriented norms in the same sort of way that beliefs are. Thus the analogy between pleasure and belief breaks down right at the point where it needs to hold firm.

This skeptical response to F1a/G2 echoes Protarchus' doubts about both the Grounding Thesis and the analogy between pleasure and belief. Protarchus agrees with the conditional statement that "if a pleasure or a pain is mistaken about what it is pleased or pained at," then it is bad (37e5-9), but he implicitly rejects the antecedent. In his view, pleasures in themselves cannot make such mistakes.[86] When Socrates suggests that pleasures often "arise with" false beliefs (37e10-11), Protarchus' reply is firm: "in such a case, we say that the *belief* is false; no one would ever assert that the *pleasure itself* is false" (37e12-38a2). As we have seen, Protarchus winds up abandoning this claim when he comes to accept both that pleasures can have false content, and that this is a way for pleasures to be false. But Protarchus need not have phrased his objection in this way. He could have said that if a pleasure "arises with" a false belief, then the pleasure is false—in the sense that it has false content—but *only the belief is mistaken*. On a view of this sort, pleasures are unlike beliefs in that they do not have an independent representational role, and so do not carry commitments to the truth of their contents. Thus agents *qua* believers can make mistakes about how things are, but agents *qua* pleasure-takers cannot. Developed in this way, Protarchus' objection amounts to the claim that the falsehood of an agent's pleasure simply cannot explain its being bad, even (and especially) if that pleasure owes its content to one of the agent's beliefs.

2.11. A Fresh Start

The case for F1/GT has finally unraveled. Neither F1 nor F1a provides an account of false pleasure powerful enough to support even the weakened version of Plato's thesis.[87] In order to preserve GT, then, Plato needs to abandon his first account of false pleasure and develop a different one. A successful account, I take it, needs to accomplish three main objectives. First, it must assign to pleasure a truth-oriented representational role. Claiming that pleasures are mistaken (and hence bad) because they are false makes sense only if pleasures have some role in representing how things in fact are. Second, it must distinguish the representational role of pleasure from that of belief, since otherwise there is no principled reason to deny that pleasures *just are* beliefs. Third, it must cast the representational role of pleasure in such a way that a pleasure's representational failure plausibly explains why the tribunal of ethical intuition condemns it as bad. In sum, then, Plato needs to develop an account that yields to all pleasure-attitudes an intrinsic representational role that is unique to them, such that their failure in this role plausibly explains their suspect ethical status.

---

[86] According to Gosling and Taylor (1982, 431-432) and Moordian (1996, 93, 96), Protarchus holds not only that no pleasures can be false, but also that all pleasures must be true. This may indeed be Protarchus' view, but his objection can be developed in a different and more effective direction.

[87] I agree with Gosling and Taylor (1982, 452) when they suggest that the main account of false pleasure does not "provide an adequate justification of the attribution of falsehood to pleasure, as distinct from its attribution to belief." As we shall see, however, I suspect that Plato has the resources to provide such a justification.

The most straightforward way for Plato to accomplish this, I think, is to claim that pleasures have the role of representing properties or states of affairs *that are good*. On a view of this sort, a pleasure is mistaken just in case it picks out properties or states of affairs that do not have the right kind of value—the kind that merits, invites, or demands hedonic appreciation. The deeper rationale for this sort of account is articulated nicely by Lovibond (1990, 215), who suggests that pleasure can be understood as a kind of *value-perception*: "The development of the capacity for value-perception just is, from one point of view, the elaboration and remodeling of the capacity for pleasure: to learn to appreciate something is to become disposed to be moved to pleasure by it."[88] Given that there are some things, properties, or states that call for this sort of hedonic appreciation, such things, properties, and states may be said to belong to (and exhaust) the set of *pleasure's proper objects*. Consequently, any pleasure taken in something other than a proper object is in error, and hence bad.[89] An account of this sort combines a value-representationalist view about pleasure with a realist (or quasi-realist) view about value. It posits both that there are true answers to questions about what has value, and that

[88] Penner (1970, 171-172 and note 7) discusses in detail the possibility that Plato thinks of pleasure as a conventional kind of perception. But the idea of pleasure as a distinctive kind of *value*-perception does not seem to occur to him.

[89] In sections 184 and 186 of his commentary on the *Philebus*, the Neoplatonic philosopher Damascius suggests, but does not argue, that something close to this is actually Plato's view. According to Damascius, there are two primary ways Plato recognizes for a pleasure to be false: either the object of pleasure (τὸ ἥδον) is "not present" (παρεῖναι) or it is "not good." If the object is not good, then the pleasure is false because pleasure itself *aims at* the good: "the love of the ugly, or of what is not beautiful, is false, since it is a striving that seeks the beautiful" (ψεῦδος γὰρ τῆς φιλοκάλου ὀρέξεως ὁ τοῦ αἰσχροῦ ἔρως ἢ ὁ τοῦ μὴ ὄντος καλοῦ). See Westerlink (1959, 86-88). Gosling (1959, 50) seems to discern that Plato's account *should* take this form: "for any given state of affairs there is an appropriate reaction, and just as when we make a judgment we are aiming at the right judgment about the thing, so when we react we are aiming at the right reaction." But Gosling goes on to say that this "cannot be what Plato has in mind" (50). As we shall see, however, there are good reasons not to share Gosling's pessimism.

pleasure-taking aims to deliver to the pleasure-taker only those answers that are true. One who holds this position has the power to claim not only that there are clear correctness-conditions that apply to each propositional pleasure, but also—and most importantly—that these correctness-conditions are determined by the *truth*-conditions of a specific propositional content.

Many questions can of course be raised about this sort of account. But before delving into these questions, I want to argue that Plato himself ultimately comes around to it—or at least something very close to it.

2.12. A New Account

In the latter half of the *Philebus* Plato advances an argument that I believe hinges on the idea that pleasures are uniquely sensitive to value. This argument, taken as a whole, runs from 53c4-55a11; but the part that bears directly on the problem of false pleasure begins at 54e1. At this stage of the argument, Socrates has just concluded that pleasures cannot belong to the class of final goods—those things that are worth choosing for their own sake—since every pleasure (according to the *process theory*) is a "process of becoming" (γένεσις). His reasoning, roughly, is that no process of becoming can be an end, and that only ends can be final goods; therefore no pleasure can be a final good.[90] This conclusion is then carried over into the second part of the argument.

[90] For more detailed analysis and discussion of this argument, see chapter 3.

Socrates frames the second part of the argument with an interesting rhetorical trope. He claims that the esteemed author of the process theory will laugh, both at those who believe that pleasure is a final good and at those who take pleasure in being pleased (54d6-e2). The talk of laughter here, which might seem at first to add nothing substantive to the discussion, is in fact highly significant when understood in the broader context of the dialogue. Just a bit earlier in the discussion (48a-50d), Socrates embarks on a lengthy, technical, and somewhat bizarre discourse on the topic of "the ridiculous" (τὸ γελοῖον). His central claim is that a person is ridiculous just in case he is mistaken about how well he is doing—e.g., in looks, wealth, or virtue (48c4-49a2). So when Socrates frames the transition to the second part of the argument with talk of laughter, we are invited to recall this earlier claim, and to seek out signs of error in those who believe that pleasures are good in themselves and take pleasure in being pleased.

> Socrates: … it is clear that [the author of the process theory] will laugh at those who claim that pleasure is good [τῶν φασκόντων ἡδονὴν ἀγαθὸν εἶναι καταγελᾷ].
> Protarchus: Most definitely.
> Socrates: And this same man [i.e., the author of the process theory] will also laugh at those who take processes of becoming as their end [καὶ τῶν ἐν ταῖς γενέσεσιν ἀποτελουμένων καταγελάσεται].[91]
> Protarchus: Why? And what sort of people do you mean?

[91] The sense of τῶν ... ἀποτελουμένων is somewhat difficult to gauge. As Bury (1897, 126) notes, ἀποτελουμένων can be taken as a neuter, in which case the objects of ridicule would be the processes associated with satisfying desires such as hunger or thirst. But then the ὅσοι at 54e4 would have to be emended. Better to take ἀποτελουμένων as masculine in the middle voice, as I have above. As for the verb itself, ἀποτελέω in the active typically means "to complete" or "to bring to an end" (cf. *Rep*. 443b7), and Plato uses it in the *Gorgias* to mean "to satisfy" or "to fulfill" one's appetites or desires (αἳ μὲν τῶν ἐπιθυμιῶν ... ταύτας μὲν ἀποτελεῖν, 503c7-d1). Since ἀποτελέω seems broadly to connote end-directed activity, I have translated the phrase as "those who take processes of becoming as their end." Cf. Hackforth (1972, 107).

> Socrates: I mean those who, when curing hunger or thirst or anything that a process of becoming cures [ὅσα γένεσις ἐξιᾶται], rejoice on account of the process of becoming insofar as it is a pleasure [χαίρουσι διὰ τὴν γένεσιν ἅτε ἡδονῆς οὔσης αὐτῆς] and claim that they would not agree to live without thirsting and hungering and experiencing all the effects that follow upon [thirsting and hungering] (54e1-6).

Socrates clearly thinks that these people—I will call them *pleasure-lovers*—are making some sort of serious mistake. He mentions two distinct episodes of pleasure in the passage: one is the curative process itself, and the other is the pleasure the agent takes *in* this process *insofar as it is* a pleasure. Since merely undergoing a curative process seems unobjectionable, even to Socrates, it must be the pleasure-lover's *second-order* pleasure that he thinks is mistaken.[92] His description of this second-order pleasure repays close attention, because it throws into relief the inadequacy of Plato's first account of false pleasure. According to that account, remember, a pleasure taken in x *as* F is false just in case it is false that x is F.[93] Now if pleasure-lovers take pleasure in restorative processes *as* pleasures, and if these restorative processes *are* pleasures—as Plato's own preferred theory has it—then the pleasure-lover's second-order pleasures are not false but *true* under F1. What this suggests, of course, is that F1 is not up to the task of pinpointing the source of the pleasure-lover's error *as Plato himself sees it*.

---

[92] Normally διά with the accusative and ἅτε with the genitive absolute have *causal* force. So Gosling (1975, 56) translates: "… are delighted because of the relevant process, because it is a pleasure." But the "because" here must be of the sort we find in the sentence "I am pleased because I have won the lottery," where it serves to pick out the content of the speaker's pleasure, rather than its cause. If we take the "because" in Gosling's translation as strictly causal, then no sense can be made of the idea that the pleasure-lovers are committing any absurd error here: they are merely undergoing curative processes. Frede's (1993, 65-66) translation ("… take delight in generation as a pleasure") is better, since it emphasizes that the pleasure-lover takes pleasure *in* curative processes *as* pleasures.

The wider context of the argument indicates that the pleasure-lover's error is evaluative in nature. After all, the argument as a whole aims to settle the question whether any pleasures can be final goods, and the first part of the argument establishes a negative answer to this question. So it makes good sense to suppose that the pleasure-lovers described in this second part of the argument are committing the same error that the first part *reveals as* an error, namely, that of taking pleasures to be final goods. It might be thought that this error is just the pleasure-lover's *believing* that pleasures are final goods. But it is not at all obvious that this is what Socrates thinks. At the end of the first part of the argument, he claims that the author of the process theory laughs at "those who claim that pleasure is good" (τῶν φασκόντων ἡδονὴν ἀγαθὸν, 54d6-7). Yet at the beginning of the second part he claims that the same person *also* laughs at "those who find their end in processes of becoming" (καὶ τῶν ἐν ταῖς γενέσεσιν ἀποτελουμένων, 54e1-2). There is an important distinction lurking here, I think. For it is one thing to attribute error to people who make a certain claim about the value of pleasure; it may be quite another to attribute error to people who "find their end" in experiencing pleasure. The former clearly targets an error of belief—a doxastic error—but the latter need not. Everything depends on what Socrates means when he says that pleasure-lovers "find their end" in processes of becoming.

When Protarchus directly *asks* Socrates what he means by this puzzling phrase, Socrates replies that by those who "find their end" in processes of becoming he means

[93] For arguments that pleasure can be taken in an object or activity *under a description*—and can thereby inherit propositional content without being taken in propositions *per se*—see section 2.2. Cf. also Frede

those who (A) "rejoice" [χαίρουσι] in processes of becoming insofar as they are pleasures, and (B) "claim" [φασι] that they would not agree to live without desire and everything that follows upon it. Now (B) certainly describes a doxastic error, but things are not so clear with (A). At no point in the dialogue does Socrates suggest that taking pleasure in something is just a way of believing something, so it is at least possible that (A) does not describe a case of believing anything at all. And yet, as we have seen, Socrates does suggest that (A) describes an error—indeed the primary error, since the logic of the passage suggests that the error in (B) is driven, at least in part, by the error in (A). It is the pleasure-lover's second-order pleasure that seems to ground his unwillingness to live without first-order pleasure. So it remains something of an open question whether the error in (A) is supposed to be a matter of believing something or not.

In any case, Socrates is quite clear that he takes the error in (A) to be "ridiculous." And as we have already seen, Socrates claims that people are ridiculous just in case they are mistaken about how well they are doing. Since it is safe to assume that for Plato agents really are doing well just to the extent that they have acquired genuine final goods, it makes sense to suppose that he thinks of pleasure-lovers as "ridiculous" in a fairly straightforward sense: they wrongly take it (in some way) that pleasures are final goods. Yet this suggests that for Plato the functional role of pleasure-taking is to *pick out* final goods. Otherwise it would make little sense for him to imply that pleasure-lovers make a mistake about what is genuinely valuable when they take pleasure in pleasure itself. What

---

(1985, 178-179) and Cooper (1999, 161-162).

Plato thinks pleasure-taking is *for* is revealed by how he thinks it *goes wrong*. And as we have seen, he supposes that pleasure-taking goes wrong when it is directed at objects that do not have the requisite final value. One of these objects, in his view, is pleasure itself. But his general account need not be constrained by this. The argument he puts in the mouth of Socrates implies that pleasure-taking in general is a mode of *appreciating what has value*, and this mode of appreciation can get things wrong even when its object is not a pleasure as such.

2.13. Virtues of the New Account

If Plato's claim is that a pleasure is false just in case it is taken in something that is not a final good, then I believe he stands a chance of vindicating GT—or at least a suitably tailored version of it. Plato's first account of false pleasure (F1) proves inadequate because it holds that the ethical status of each pleasure is determined exclusively by the truth-value of its content. As we have seen, this account gets him nowhere. The new account, however, gives him room to maneuver. For if he accepts it, he is entitled to claim that the ethical status of a pleasure depends in no way upon the truth-value of its content. Indeed, he *must* claim this. When one is pleased that *p*, the truth of <*p*> is irrelevant to the issue of whether genuine value would be realized if <*p*> were true.[94] Yet this is precisely the issue that, according to the new account, determines whether the

pleasure is true or false. From this it seems to follow that the content whose truth each pleasure is committed to is not equivalent to that pleasure's *own* content. Each pleasure's own content must be modified to yield the uniquely value-oriented content whose truth that pleasure is committed to. For what makes a pleasure mistaken, on this account, is not the falsehood of its content, but the failure of its content to pick out a state of affairs (or a property) whose realization (or instantiation) would be genuinely valuable.

Needless to say, Plato does not develop this new account in any detail within the body of the *Philebus*. But it is worth trying to formulate the account more precisely so as to get a firmer grasp of its power. Pleasures that are taken in states of affairs (e.g., being pleased that it is raining) can be given the following analysis:

> (F2) For any proposition <p>, let <q_p> be the proposition that it would be good if <p> were true. Then, for any pleasure H with content <p>, H is false just in case <q_p> is false.

A similar account can be given for pleasures that pick out properties rather than states (e.g., taking pleasure in smoking as a form of rebellion):

> (F2) For any object x and any property F, let <r_{Fx}> be the proposition that x is good insofar as it is F. Then, for any pleasure H in x as F, H is false just in case <r_{Fx}> is false.

If Plato enlists F2 in his attempt to defend GT, then he needs to defend the following two claims:

---

[94] Gallop (1960, 332) makes this important point in a slightly different way. When one hears some news, he writes, "whether the news be true or false, the pleasure one takes in it is not to one's discredit if it is in

(F2/GT) For any pleasure H with propositional content <p> [or: in x as F],

(F2/G1) if H is bad then <q_p> [or: <r_{Fx}>] is false, and
(F2/G2) if <q_p> [or: <r_{Fx}>] is false then H is bad *because* <q_p> [or: <r_{Fx}>] is false.

F2/GT is poised to succeed where F1/GT fails. For F2/GT is not vulnerable to the sort of objections that dispose of F1/G1 straightaway and F1/G2 more tortuously. As we saw when examining F1/G1, what makes Schadenfreude and the intense bodily pleasures bad is certainly not their having false content, since only some of them are false in this sense. What seems to make Schadenfreude bad is that it is a pleasure whose content represents a state of affairs (someone's suffering a misfortune) that is a bearer of significant disvalue. F2/GT captures this intuition nicely.[95] Likewise, Plato's disapproval of intense bodily pleasures is better captured by F2/G1 than by F1/G1. Though he seems to think that such pleasures have bad consequences for the agents who have them, he bases his critique on the idea that these pleasures are taken in mere bodily processes—events that cannot, in his view, be valuable in themselves. Without pronouncing at the moment on the merits of Plato's largely dismissive view of bodily pleasures, we can at least see that F2/GT is far better than F1/GT at making his view both intelligible and plausible.

---

fact the reaction that *would* be appropriate if the news were true."

[95] Cf. Irwin (1995, 320): "A hedonist might say that the cruel person's pleasure is one good feature of the situation, however much the goodness may be outweighed by bad features of the situation. According to Plato, however, the cruel person's enjoyment is another bad feature."

As F2/GT resolves the problem of Schadenfreude, so it resolves the problem of fantasy. Whereas F1/GT implies that that there is something inherently wrong with the pleasures of counterfactual imagination, F2/GT does not. What would make such a pleasure wrong, according to F2/GT, is not its having counterfactual content, but rather its having content that expresses a state of affairs or a property the realization or instantiation of which would not be a final good. One might object that even this is too tyrannical, on the ground that there is nothing inherently wrong with the pleasures of fantasy, no matter what their content might be. But the assumption driving this objection is at least debatable. Propositional pleasure does seem to be a mode of appreciation, and it is not too farfetched to suppose that there is something wrong with appreciating what has disvalue—even if this appreciation does not involve any commitment to the *realization* of such disvalue. At any rate, this view is certainly preferable to one that condemns all the pleasures of fantasy simply on the ground that they have false content.

So F2/GT is far better than F1/GT at reflecting our (and Plato's) ethical intuitions about the familiar problem cases. More important, however, it provides ground for a firm distinction between *hedonic* error and *doxastic* error. According to F2/GT, the pleasure with content <*p*> and the belief that it would be good if <*p*> were true have the same truth-based correctness-conditions. But it does not follow from this that they are the same sort of attitude. Indeed, there is an important difference between them. What makes a belief correct is just what makes its content true; but, as we have seen, the correctness of a pleasure is not determined in this way. What makes a *pleasure* correct is what makes

it true that the property or state of affairs expressed by its content is *good*. However, and despite their difference in this respect, pleasure and belief both have correctness-conditions that are grounded in truth-conditions. Pleasure has a representational role, albeit one that is sensitive to value in a way that belief is not. It aims intrinsically, exclusively, and uniquely to get the *world of value* right.

This distinction between hedonic and doxastic attitudes is of course too bloodless to capture the *felt* difference between believing and pleasure-taking. But it can accommodate this difference, and can do so in a way that makes interesting sense of prickly mental phenomena such as hot akrasia. Since pleasures carry commitments to the truth of certain contents, and since these contents can stand in rational relations with the contents of other mental states (such as beliefs and desires), pleasures can play a direct role in rational deliberation and the justification of action. But because pleasures are not beliefs, it is possible for a rational agent to be led by hedonic considerations to perform an action A without ever believing that A-ing would be good to do, and indeed, without ever *abandoning* the belief that A-ing would be *bad* to do. To understand pleasure in this way is to see it as a mode of appreciation that has its grip on us whether or not we give our doxastic assent to its (modified) content. This at least approximates the evident difference between what it's like to believe something and what it's like to be pleased about something, and does so in a way that makes good intuitive sense of the possibility of rational akrasia.

To sum up, then, F2 is an account of false pleasure that yields to all pleasure-attitudes an intrinsic representational role unique to them as such, and that, in combination with GT, does an admirable job of showing how their failure in this role explains our inclination to classify them as ethically suspect. Given the support of F2, Plato's Grounding Thesis is a coherent, philosophically interesting view with some surprising and attractive features. As we shall see in the next chapter, this view about the ethical status of pleasure and pain supports Plato's anti-hedonism in a subtle and intriguing way.

3. Plato's Anti-Hedonism

3.1. Introduction

It is widely agreed among commentators that the *Philebus* constitutes, at least in part, a sustained attack on ethical hedonism. But this consensus survives only at a very general level. The structure of Plato's attack, the nature of its target, and the cogency of its results are all matters of continuing debate. One of the many unresolved questions in this debate concerns Plato's view about the *final value* of pleasure. Does he hold that at least some pleasures are worth pursuing for their own sake? Or does he hold that no pleasure, regardless of its content or consequences, is anything more than derivatively valuable? Most interpreters have argued in favor of the first alternative, claiming on Plato's behalf that a pleasureless life is missing something of final value just insofar as it is pleasureless.[96] On this reading—which I will call the *liberal view*—Plato wants to include in the class of final goods those pleasures that have earned his approval by being taken in the right sort of things or activities. My aim in this chapter is to show that the liberal view is wrong. Plato denies that any pleasure is to be valued for its own sake, even if its content is noble and its consequences harmless. For his theory holds that pleasures *as such* have value only insofar as they make possible the achievement of non-hedonic

[96] Cf. Irwin (1995, 332-337), Cooper (1999, 150-164), and Carone (2000, 257-264). Dissenting is Frede (1992, 440), (1993, lvi).

ends. As I hope to show, the central plank of Plato's anti-hedonist platform is the claim that we choose incorrectly when we choose to do something for the sake of the pleasure it affords—even if this is not our *only* reason for so choosing. Though this doctrine might seem to make Plato's ethical theory intolerably austere, I will suggest that the reasoning behind it is deeper, more challenging, and ultimately more attractive than one might otherwise expect.

The debate over Plato's view of pleasure's value in the *Philebus* turns on the interpretation of two separate arguments, both of which feature prominently in the body of the dialogue. The first of these I call the *choice argument* (CA) (21d6-22b8; cf. 60b7-61a2); the second I call the *process argument* (PA) (53c4-55a11). While the CA concludes that the most choiceworthy human life is neither wholly unintellectual nor wholly pleasureless (22b3-8), the PA concludes that no pleasure can belong to the class of things that are worth pursuing for their own sake (55a9-11). On the face of it, these conclusions are compatible. But trouble arises when we try to figure out *why* Socrates thinks that a life with some pleasure in it is more choiceworthy than a life with maximal intelligence and no pleasure. According to the liberal view of the CA, Plato's reasoning here hinges crucially on the implicit claim that a pleasureless life cannot be most choiceworthy for a human being *precisely because* it is pleasureless. The disqualifying feature of this life is just its lack of pleasure, rather than its lack of some other sort of thing that, for whatever reason, cannot be included in the life unless pleasure is. The problem with attributing this claim to Plato, however, is that it flatly contradicts the

conclusion of the PA. Plato cannot maintain at once that some pleasures are final goods (by the CA) and that no pleasures are final goods (by the PA). None of the advocates of the liberal view address this apparent contradiction, perhaps because they are not convinced that Plato wants the PA to apply to all pleasures equally.[97] In their view, the conclusion of the PA can be effectively explained away as a doctrine that Plato does not hold in its most general form.

After examining the PA in detail, however, I argue that there are no compelling reasons to think that Plato finds the argument unsound. I then attempt to dissolve the apparent contradiction between the PA and the CA by showing that the liberal interpretation of the CA is forced. The CA need not be read as establishing that some pleasures are final goods; the argument goes through on the strength of a much weaker assumption, namely, that some pleasures are *indispensable derivative* goods. For if some pleasures are indispensable features of the good life, then a wholly pleasureless life cannot be the most choiceworthy. It is missing something of final value, to be sure, but this is only because it is missing something (other than pleasure-taking) the acquisition of which *requires* pleasure-taking. Since this construal of the CA is compatible with the evident conclusion of the PA, I argue that the liberal view should be rejected.

[97] See, for example, Gosling and Taylor (1982, 153-154) and Carone (2000, 264-270).

3.2. The Process Theory

As an anti-hedonist dialogue, the *Philebus* reaches its climax in the process argument. For this is the first argument to persuade Protarchus—the dialogue's appointed defender of hedonism—that no pleasure of any kind can belong to the set of good things (55a9-11). Since Protarchus cannot both concede this and maintain any recognizable version of his original hedonism, his unambiguous assent to the conclusion of the PA marks both his own final defeat and, for this reason, the structural climax of the dialogue as a whole.

But the argument itself emerges rather suddenly in the text, and the very suddenness of its appearance has led some commentators to argue that it is at best an afterthought, and at worst a half-hearted parting shot.[98] This judgment is too hasty, however. Socrates presents the argument directly after giving an extended account of the pure pleasures (50e5-53c2). His standing suggestion at this point is that these pleasures (unlike the impure kind) might be worth cultivating after all (53b8-c2). So the sudden appearance of the PA at this point in the dialogue makes structural sense if its purpose is to modify or refine this standing suggestion. Plato's decision to place the PA here (rather than somewhere else) is perfectly justified if part of the point of the argument is to insist that, no matter how beneficial some types of pleasure are, *even these types* bear a certain

[98] The leading representative of this view is Hackforth (1972, 105-6), who proposes that we read the PA as a hastily constructed appendix. Likewise Gosling (1975, 220-1).

generic property that renders them unworthy of ultimate pursuit. As we shall see, I believe that this is in fact the point of the PA. But my reasons for thinking so will become clear only once we have examined the argument itself in greater detail.

The crucial opening gambit of the PA is an initially obscure proposal about the nature of pleasure, which I will call the *process theory* (PT). Socrates presents the theory as follows:

> Have we not heard about pleasure that it is in every case a process of becoming, and that no pleasure is in any way a being? [περὶ ἡδονῆς … ὡς ἀεὶ γένεσίς ἐστιν, οὐσία δὲ οὐκ ἔστι τὸ παράπαν ἡδονῆς] For certain elegant minds [κομψοὶ γὰρ δή τινες] attempt to reveal this account to us, and we should be grateful to them. (53c4-7)

Protarchus finds this theory puzzling at first (53c8) and so should we. From this extremely compressed account alone it is nearly impossible to see what the central distinction means, how its two terms are related, and why it is relevant to the analysis of pleasure in the first place. Socrates seems to acknowledge this (53c9-d1), and proceeds to illuminate the original distinction with four examples, each of which is plainly supposed to apply to the two original terms (i.e., being and becoming):

(1) A being is something "itself by itself" [αὐτὸ καθ' αὑτό] while a process of becoming "aims in every case at something else" [ἀεὶ ἐφιέμενον ἄλλου]. (53d3-4)

(2) Beings are "in every case most venerable" [σεμνότατον ἀεὶ] while processes of becoming are not. (53d6-7)

(3) Beings are to processes of becoming as "good-looking and upstanding boys" [παιδικά καλὰ καὶ ἀγαθὰ] are to "their manly lovers" [ἀνδρείους ἐραστὰς αὐτῶν]. (53d9-10)

(4) Processes of becoming are "in every case for the sake of some one of the beings" [ἕνεκά του τῶν ὄντων ἔστ᾽ ἀεί] while beings are "in each case that for the sake of which what becomes for the sake of something in every case becomes" [οὗ χάριν ἑκάστοτε τὸ τινὸς ἕνεκα γιγνόμενον ἀεὶ γίγνεται]. (53e4-7)

Taken together, these four examples suggest a coherent view about the nature of pleasure. Though elliptical, (1) suggests that beings are self-sufficient in a way that processes of becoming are not, since the latter, unlike the former, constitutively "aim at" things other than themselves. And while (2) suggests that what is aimed at belongs to a higher order of value than what aims, (3) implies that this higher order of value explains their being aimed at in the first place. The "good-looking and upstanding boys" are the targets of "manly love" presumably because they have a certain value to which the lovers are attracted. In (4) Socrates introduces the "for the sake of" relation to capture the sense in which processes of becoming "aim at" beings. At this point Protarchus turns petulant—perhaps because he thinks that Socrates is being unnecessarily opaque—and suggests a gloss on (4) that exploits a familiar craft-analogy. As Protarchus sees it, the issue is "whether shipbuilding is for the sake of ships or ships for the sake of shipbuilding" (54b2-4). Socrates concurs, saying "that's just what I mean" [λέγω

τοῦτ' αὐτό] (54b5).[99] He then proceeds to give a more extended account of what all of this amounts to:

> I hold that all ingredients, as well as all tools, and quite generally all materials, are always provided for the sake of some process of becoming. I further hold that every process of becoming in turn always takes place for the sake of some particular being, and that all becoming taken together takes place for the sake of being as a whole. (54c1-4)

Here Socrates seems to extend the proposed craft-analogy to embrace three different types of thing, all of which belong to a general hierarchical structure. His suggestion is that tools, ingredients, and materials are "for the sake of" processes of becoming, and processes of becoming are "for the sake of" beings. And by this he seems to mean that the function or purpose of each of the subordinate items is to realize or acquire one of the superior items. Take the example of shipbuilding: the ingredients and tools required to build a ship (e.g., wood, metal, saw, straightedge, etc.) have as one of their (potential) functions the production of ships—while this very act of production aims at, and hence realizes if unimpeded, a ship.[100]

In light of this elaboration, the four examples provided by Socrates seem to support the idea that a "process of becoming" is *essentially teleological*. Something's being a "process of becoming" depends on its aiming at some particular end, where this

[99] *Contra* Tuozzo (1996, 503), who wrongly claims that "Socrates never explicitly endorses the view that pleasure is a "coming-into-being" in the way that shipbuilding is." It is important to see, in any case, that the craft-analogy must not be taken as a substantive claim about the nature of all processes of becoming. Surely neither Protarchus nor Socrates means to say that every process of becoming is a case of technical production. At best, this gloss suggests only the converse: every craft-production is a process of becoming.
[100] For further discussion of the implications of PT and the craft-analogy for Plato's later ontology, see Shiner (1974, 49-105) and Benitez (1989, 103-105).

end is something other than itself. Every "process of becoming" is in this sense *teleologically subordinate*. A "being," on the other hand, is self-sufficient; it is not teleologically subordinate to any end, and it is *teleologically superior* to every process of becoming. According to Socrates, then, processes of becoming take beings as their constitutive ends. Shipbuilding, in order to be what it is, must be directed toward the production of a ship, where the ship itself is something distinct from shipbuilding. Notice that the issue here is not whether one can engage in shipbuilding for its own sake—obviously one can—but whether one can engage in shipbuilding without (in some sense) aiming to produce a ship. It is only this latter claim that Socrates denies. His view seems to be that every process of becoming must aim at some end other than itself, where this end is just whatever it is that the process in question hits upon when it is correctly completed.

It is worth asking at this point how the distinction between beings and processes is supposed to help us understand what pleasure is. If my reading of the distinction is correct, then Socrates holds that every pleasure is a process that constitutively aims at some state of being that has a higher order of value than the process itself. At first blush, however, this claim seems opaque. What sense does it make to say that every pleasure is a process of becoming? Assuming that sense can be made of this claim, what is the end state of a pleasure? Socrates does not explicitly answer any of these questions in or around the section of the *Philebus* in which the PA appears. But if the interpretation I

have offered in the previous two chapters is correct, then Plato's overall account of pleasure provides two interesting and interrelated sorts of answer.

The first sort of answer emerges from considerations I developed in chapter 1. There I argued that Plato deploys a model of pleasure—the *equilibrium model* (EM)—according to which every pleasure is a process of restoring a state of psychological harmony. It is not difficult to see how EM and PT might be related. Indeed, on the assumption that every "process of restoring" is also a "process of becoming," EM *implies* PT.[101] Among other things, EM holds that the end states of restorative processes are, unlike these processes themselves, "beings." As Socrates puts the point, pleasure in general is, for the organisms who undergo it, "the path to their being" (τὴν δ' εἰς τὴν αὐτῶν οὐσίαν ὁδόν, 32b3). On any reasonable construal of EM, then, restorative processes (*qua* restorative) *aim at* an equilibrium state of "being," since undergoing them is just the way for an organism to return to this state. So from either EM or PT we get the result that there is a state of being "for the sake of which" every pleasure happens. Unlike PT, however, EM specifies this state as a state of equilibrium, and yields a fairly elaborate account of why it makes sense to call the process of restoring this state a pleasure.[102]

The second and more interesting sort of answer springs from the idea, which I attribute to Plato in chapter 2, that pleasures are on a par with beliefs in that they are attitudes evaluable on the basis of their content. Plato's analogy between pleasure and

[101] Cf. Gosling and Taylor (1982, 153) and Tuozzo (1996, 503).

belief implies that pleasures are directed toward a content the value of which determines their own value. Just as the value of a belief-attitude (as correct or mistaken) is given by the truth-value of its content, the value of a pleasure-attitude is given by the *goodness*-value of its content. Now the correctness-conditions of either attitude can perhaps best be understood as conditions on the sort of content that the attitude in question is intrinsically aiming to have. When a belief-attitude has false content, for example, it is mistaken precisely because it intrinsically aims (*qua* belief) at having *true* content.[103] Thus the mistake itself is cashed out as a failure to strike the relevant target. Given that the same line of reasoning holds for pleasure-attitudes as well, both pleasures and beliefs can be plausibly construed as teleologically subordinate to that at which they aim.

Though these two answers are very different in form, they are both well-entrenched in the *Philebus*, and for this reason alone I think it is appropriate to draw on both of them when interpreting the PA. As we shall see, they illuminate the argument in striking and unexpected ways.

3.3. The Process Argument

The importance of the process theory as a premise in the process argument should by now be obvious. For if every pleasure is teleologically subordinate to that at which it

[102] Tuozzo (1996, 504) resists the idea that the process in question here needs to be understood as a restoration of equilibrium, but he thinks the PA goes through anyway.
[103] See *Theaetetus* 193e6-194a4, where Socrates likens false belief to a misfired arrow: "like a bad archer, he shoots far of the mark and misses [ἁμαρτεῖν], and this, then, is what is called false." For illuminating discussion of this and similar passages, see Caston (1993, 226-230).

aims, then no pleasure can have the same kind of value that its end does. Socrates exploits this thought by claiming that the membership conditions for the "class of the good" exclude anything that intrinsically aims at something other than itself. The core of his argument is as follows:

(1) Every pleasure is a process of becoming. (53c4-5; 54c6-7) (PT)

(2) The end of every process of becoming is "in the class [μοίρᾳ] of the good, but that which becomes for the sake of something else must be placed in another class." (54c9-11; cf. 54d1-2)

Therefore,

(3) "Those who claim that pleasure is good" [τῶν φασκόντων ἡδονὴν ἀγαθὸν εἶναι] are wrong. (54d6-7) No pleasure is good. (54e1-2)

The conclusion here is radical. For Socrates is denying not only that pleasure is *the* good, but that *any* pleasure is good *at all*.[104] The primary supporting claim for this conclusion is premise (2), which asserts that no process of becoming can be a good. This claim should strike us as puzzling, however, since it does not seem to fit very well with the theory that Socrates has just elucidated. For if the end state of every process of becoming has final value, then presumably every process of becoming has value too—even if this value is somehow inferior to that of the end state. It seems unreasonable for Socrates to deny that processes of becoming have any value at all since he also wants

[104] *Contra* Carone (2000, 266), who seems to think that the argument is targeted at the claim that pleasure is the *only* good. As even a cursory glance at the conclusion of the argument will show, the text will not bear that interpretation.

to say that these processes bring about or realize value. But then why does he insist that pleasures cannot be good?

The best way to make sense of this claim, I think, is to suppose that by the "class of the good" he means the set of *final* goods. It is appropriate for Socrates to exclude processes of becoming from this set, since—on his view—the value of these processes intrinsically depends upon the value of the beings at which they aim. As a result, they have value only in relation to other things, and not in themselves. If this interpretation is right, then the conclusion of the PA jibes nicely with the theory that motivates it. Since every pleasure is essentially a teleological process—and therefore not an end in itself—no pleasure can be a final good. So the upshot of the PA as a whole is perhaps less radical than a cursory look at its conclusion might suggest. In denying that any pleasure is good, Socrates is not claiming that all pleasures are worthless. Indeed, as we shall see, this is pretty nearly the opposite of what he thinks. All he is insisting upon here is that pleasures, as such, cannot be bearers of *final* value.[105]

But is even this conclusion warranted? As I have reconstructed it, the PA turns on the potentially dubious assumption that no teleological process can be an end in itself. In the *Republic* Plato makes conceptual room for things that are good both for the sake of their consequences and for their own sake (357c). Indeed, in that dialogue he holds out "harmless pleasures" as examples of things that are good for their own sake alone (357b).

---

[105] Gosling and Taylor (1982,153-4) rightly claim that the process theory applies to all and only those pleasures which can be analyzed under EM. But on their view, this restriction spares both the anticipatory pleasures and the pleasures of virtue from the conclusion of the PA. I disagree. As I argued in chapter 1, EM applies to *every* type of pleasure, including those mentioned by Gosling and Taylor. So, *mutatis mutandis*, no type of pleasure escapes the PA. Cf. Frede (1993, lv).

Why the sudden stinginess? Perhaps the difference between these two positions rests upon Plato's fuller appreciation, in the *Philebus*, of the consequences of EM for his value theory. In this dialogue, perhaps, he first comes to think that things can be final goods only if they do not constitutively aim at something else. The problem with this claim, even if Plato does hold it, is that the pleasure-lover might simply deny it. He might concede that pleasures are restorative processes, but nonetheless assert that he pursues these processes for their own sake. Just as the shipbuilder might want to engage in shipbuilding regardless of how his ships turn out, the gourmet might want to undergo gustatory pleasures regardless of their consequences for his health. If these choices make sense—and they certainly seem to—then they show how one might accept EM without conceding that processes are never final goods. So even if one accepts that pleasure constitutively aims at something other than itself, one can apparently still value it as an end without lapsing into irrationality.

It is not clear how Plato could respond effectively to this sort of challenge without providing some additional argument. The PA by itself relies on the disputed premise, and so cannot be marshaled in defense of that premise. But there is a further argument to be found in this section of the dialogue, appearing just after the PA is concluded. This second argument, as I read it, is designed to show that anyone who attributes final value to a teleological process must have irrationally ordered preferences. If sound, such an argument would provide support for Plato's previous supposition that teleological processes cannot be bearers of final value.

3.4. The Argument from Irrational Preferences

As I read it, the argument that supplements the PA is a *reductio* against the claim that some pleasures are final goods (54e1-55a11). Socrates' strategy is to tease out an absurd consequence of this view by showing that anyone who holds it must assume that the same end both has and does not have final value. In the opening stages of this supplementary argument, Socrates asserts that the "elegant" author of the process theory will laugh at those who claim that any pleasure is good. This, I think, constitutes the conclusion of the first argument. But evidently another line of argument begins where the first one ends:

> Socrates: And this same man [i.e., the author of the process theory] will also laugh at those who find their end in processes of becoming [καὶ τῶν ἐν ταῖς γενέσιν ἀποτελουμένων καταγελάσεται].
> Protarchus: Why? And what sort of people do you mean?
> Socrates: I mean those who, when curing hunger or thirst or anything that a process of becoming cures [ὅσα γένεσις ἐξιᾶται], rejoice on account of the process of becoming insofar as it is a pleasure [χαίρουσι διὰ τὴν γένεσιν ἅτε ἡδονῆς οὔσης αὐτῆς] and claim that they would not agree to live without thirsting and hungering and experiencing all the effects that follow upon [thirsting and hungering].
> Protarchus: They're likely to, at any rate.
> Socrates: But wouldn't we all say that destruction [τὸ φθείρεσθαι] is the opposite of becoming [τῷ γίγνεσθαί]?
> Protarchus: Necessarily.
> Socrates: So someone choosing this life would choose destruction and becoming, but not that third life, in which there is neither being pleased nor being pained, but only thinking [φρονεῖν] in its purest possible form.
> Protarchus: Well, Socrates, it seems that a great absurdity [ἀλογία] follows if one holds that pleasure is good (54e1-55a11).

The target of Socrates' attack here is the agent who would not agree to live without pleasure, even if he were given the chance to live a wholly painless life of intelligent thought. It is important to notice, however, that the agent's refusal is grounded in his assumption that at least some pleasures are final goods.[106] According to Socrates, the pleasure-lover supposes that pleasures such as eating and drinking are good in themselves, and he rejects the pleasureless, painless life of the mind for this reason. So what Socrates is attacking here, I take it, could be the pleasure-lover's *grounds* for choosing as he does, rather than the actual choice itself. This is a delicate point of interpretation because, as we shall see, there is good evidence in the *Philebus* that Socrates would actually approve of the pleasure-lover's choice *per se*. For he explicitly and frequently claims that the life with at least some pleasure in it is more choiceworthy for us than the pleasureless, painless life of pure thought (22a-b, 60e, 66c).[107] If Socrates is consistent across the dialogue, then, he cannot be saying here that it is always irrational to prefer a pleasant life to an anaesthetic one. He must be saying that the pleasure-lover makes the right choice, *but for the wrong reason*.

On this reading, the absurdity that Socrates points out here—and that Protarchus is quick to notice—follows from the pleasure-lover's assumption that there are at least some pleasures worth pursuing for their own sake. As Socrates suggests, this thought commits the pleasure-lover to placing value on destruction, since destruction—according

[106] For arguments to this effect, see above, section 2.12.
[107] I deal with this and related issues in section 3.6 and following.

to EM, at any rate—enables pleasure. So anyone who attributes final value to any pleasure will be forced also to attribute value to whatever destructions, defects, and pains are enabling conditions for that pleasure. If EM is correct, then a life is pleasure-rich only insofar as it is destruction-rich, defect-rich, or both. This is presumably why the pleasure-lover refuses to live without "thirsting and hungering," for example, even though the *point* of "thirsting and hungering" is to make the pleasures of drinking and eating possible. If this interpretation is right, then the absurdity Protarchus suddenly notices is that the pleasure-lover, insofar as he pursues any pleasure for its own sake, must also be a lover of destruction, defect, and pain.

Though this interpretation helps us see how the argument can be enlisted in support of the PA, it leaves Socrates with what seems to be an incomplete line of reasoning. For it does not show us in a convincing way that the pleasure-lover is irrational in valuing things such as thirst and hunger. If an agent decides that some pleasure has final value, and undertakes to pursue it for its own sake, he will admittedly have to cultivate the conditions under which this experience is possible. And he will, as a consequence, have to place some value on defective states and destructive processes (given EM).[108] But this does not *by itself* seem to saddle him with any incompatible commitments. The pleasure-lover is in trouble, I take it, only if he cannot consistently justify pursuing defect, destruction, and (if necessary) pain for the sake of the pleasure they enable. Does Socrates give us any good reason to suppose that he cannot? One reason might be that the

pleasure-lover, insofar as he must accord value to both restoration and destruction, must accord value to *opposites*—and this might seem problematic. On the other hand, it is quite clear that the pleasure-lover need not accord the *same type* of value to each of the two opposites. He can hold, without contradiction, that some pleasures are bearers of final value, and that the destructions or defects that make these pleasures possible are (therefore) bearers of *derivative* value. He chooses pleasures for their own sake, and destructions for the sake of the pleasures they make possible. Once the pleasure-lover's position is sorted out in this way, it does not seem to suffer from any obvious absurdity.

Some commentators suggest that the thrust of Socrates' criticism here is not that the pleasure-lover is committed to valuing opposites, but that he is committed to living an unsatisfiable life. This is a kind of pathology, they think, though perhaps not one that involves contradictory claims or evaluations. Those who perpetually seek destruction for the sake of restoration are doomed to struggle without end, the thinking goes, and for this reason are open to the charge of absurdity. In laying out this charge, Frede (1993, lvi) even calls on the Greek figure most associated with the absurd: "people who choose a life of pleasure for its own sake condemn themselves to Sisyphean labors, to creating ever new needs to have something to fulfill." The pleasure-lover is a Sisyphean figure because he must continually cultivate destructive processes and defective states, and so—on this interpretation, at any rate—cannot ever achieve satisfaction. And from this it is supposed to follow that there is something deeply wrong with the pleasure-lover's life.

---

[108] Cf. Carone (2000, 268) "pleasures … gain intensity precisely insofar as they come along with painful lacks … therefore, anyone who chooses pleasures as *geneseis* would also be choosing their opposite, pain

Yet it remains to be seen whether this charge reflects anything more than the prejudice of the prosecutor. Taken at face value, Frede's claim is patently question-begging. For there are two ways in which one might lead an unsatisfiable life. If satisfaction amounts to having one's desires permanently quieted, then of course the pleasure-lover cannot lead a satisfied life. Yet it is difficult to see why the pleasure-lover should accept this construal of what it is to be satisfied, since it rules out in advance the mode of valuation that the pleasure-lover is proposing. He could either assert that he has no interest in being satisfied—and do so without significant penalty—or reject his attacker's account of what satisfaction is. A more neutral construal of satisfaction, and one more agreeable to the pleasure-lover, would have it that satisfaction consists simply in achieving one's (chosen) ends. If Sisyphus had held that there is nothing he would rather do than push a rock up a hill, there is no obvious reason to suppose that the life he wound up living was plagued with dissatisfaction. On this construal of what it is to be satisfied, the pleasure-lover's life seems no more unsatisfiable than any other—unless of course there is some argument in the offing to show that the pursuit of pleasure inevitably fails. But Socrates provides no such argument.

Is there a different way to understand the argument such that it exposes a genuine problem with the preferences of the pleasure-lover, and does so without begging the question against him? One way to do this, I think, is to interpret Socrates as saying not that the pleasure-lover is a Sisyphus figure, but that the pleasure-lover is committed to contradictory claims about the value of equilibrium states. On a reading of this sort,

---

and destruction."

Socrates' argument yields the conclusion that no one can treat any pleasure as a bearer of final value without betraying a failure to grasp what final value is. How would such an argument work, and on what grounds can we plausibly attribute it to Socrates?

As I read it, the argument turns on the idea that the pleasure-lover must value the destruction of a harmonious state for the sake of something that *intrinsically aims at establishing* that harmonious state. Once the pleasure-lover accepts EM, he must concede that pleasure (*qua* restoration) is teleologically subordinate to whatever equilibrium state it aims at, and that this equilibrium state is a final good. So Socrates needs to come up with a reason to think that this should prevent the pleasure-lover from placing anything more than derivative value on pleasure. The reason he comes up with, I think, is that if any pleasure is a final good, then its own end must *not* be a final good. For if its own end *were* a final good, then no destruction of that end could be anything but bad. But as we have seen, if some pleasure is a final good, then the destruction of the end at which it aims is an instrumental good. Therefore the pleasure-lover is confused. Once he has accepted EM, he cannot consistently hold that *any* pleasure is a final good. Let me try to set out this line of reasoning in greater detail.

For any pleasure P, there is some equilibrium state at which P aims E<=P, and some destruction (or defect) of that state D (E<=P). Now:

(1) P is a final good. (Premise to be contradicted)

(2) If P is a final good, then D (E<=P) is an instrumental good. (54e6-8)

(3) D (E<=P) is an instrumental good. (From 1 and 2)

(4) E<=P is a final good. (EM and PT)

(5) If E<=P is a final good, then D (E<=P) is an unrestricted bad.

(6) D (E<=P) is both an instrumental good and an unrestricted bad. (From 3, 4, and 5)

(7) Nothing is both an instrumental good and an unrestricted bad.

Therefore,

(8) P is not a final good.

Obviously the crucial premise here is (5), and it is certainly open to dispute. For one thing, (5) seems to rule out the possibility that some final goods can be acquired only at the expense of forsaking others. As the familiar case of Gaugin suggests, the argument proves too much. It should not rule out this possibility. Thus premise (5) should be limited so as to apply only to those cases in which the destructions in question are destructive of *the very equilibrium state at which the relevant restoration aims*. If the premise is restricted in this way, it does not rule out the possibility that some intrinsic goods can be realized only if others are not. The premise can thus be amended as follows:

(5*) If E<=P is a final good, then *if D (E<=P) does not enable the achievement of some end other than E<=P*, then D (E<=P) is an unrestricted bad.

Though a bit awkward, this modified version of the principle seems to get the right result. It would be very odd indeed if the only thing that makes the destruction of a given end-state valuable is that it enables the *restoration* of that very end-state. (5*) helps

to make this clear, and so bolsters the intuitive force of the argument. Restorative processes cannot be final goods, on this view, for if they were, the destructive processes (and defective states) that make them possible would be the *wrong sort* of instrumental goods. The same point can be illustrated in terms that Socrates uses in the argument. The drink-lover, he says, must hold that thirst itself is valuable for the sake of thirst-satisfaction. Yet the intrinsic aim of thirst-satisfaction is a state of equilibrium with respect to thirst. So the drink-lover must cultivate thirst for the sake of getting rid of thirst—which seems absurd. The argument can be applied in a similar but more challenging way to the case of learning (52a-c). Given that learning is a teleological process of coming-to-know, it is to be valued only for the sake of the knowledge-state at which it aims, and not for its own sake. One might be tempted to deny this, and to insist that learning is worth pursuing for its own sake as well. But if learning is to be valued for its own sake, then—according to the plausible premise (2)—ignorance and forgetting are to be valued instrumentally, for the sake of coming-to-know. Since this is clearly absurd, learning is not to be valued for its own sake.

So far I have interpreted this argument in terms of the teleological relation Plato develops via EM. But there is no compelling reason to restrict the argument's scope in this way. It can be interpreted in terms of the other teleological relation Plato discusses in the *Philebus*, namely, the relation between an attitude and the content it intrinsically aims to have. Plato's analogy between pleasure and belief implies that every pleasure has a content the value of which determines the pleasure's own value. Just as the value of a

belief-attitude (as correct or mistaken) is given by the truth-value of its content, the value of a pleasure-attitude is given by the *goodness*-value of its content. If this is one of the teleological relations at play in the argument, then a similarly absurd consequence can be drawn from the pleasure-lover's commitments. To value for its own sake the having of an attitude with content-determined correctness-conditions (such as a belief or a pleasure) is simply to misunderstand the nature of such attitudes. It would be bizarre, for example, to claim that belief-attitudes are worth pursuing for their own sake, as if their content did not matter. For a belief-attitude is something that aims intrinsically at what is true, so that its own value must be understood as contingent upon the truth-value of its content. This is why it would make no sense to advise someone, from an epistemological standpoint, to shoot for having as many beliefs as possible. And assuming that pleasure-attitudes are like belief-attitudes in this way, it would make just as little sense to advise someone, from an *ethical* standpoint, to shoot for having as many *pleasures* as possible. Attitudes of this sort simply cannot be bearers of final value.

The upshot of the argument from irrational preferences, then, is that one takes final aim at a teleological subordinate only on pain of absurdity. For to aim at that sort of thing is, in effect, to deny that the end at which that sort of thing intrinsically aims has independent value. It is of course possible for the skeptic to go back and reject the claim that pleasures are teleological processes—so as to dispute the status of their end-states as final goods—but, as we have seen, that move too carries a price. Insofar as pleasures are

to be construed as content-bearing attitudes, they are not easy to understand outside of the teleological framework Plato provides.

To be sure, the argument as I have interpreted it here cannot be read off the text directly. But then neither can the prevailing "Sisyphean" interpretation. In order to make sense of the reasoning behind what Socrates says here, some interpolation is required. At least my interpretation, unlike the Sisyphean one, does not saddle Socrates with a patently question-begging argument against the pleasure-lover. On the contrary, it credits him with an argument that is plausible—at least on the assumption that EM is roughly correct and that Plato's analogy between pleasure and belief is apt. Moreover, it illuminates a way in which the argument reinforces the directly preceding PA. The obvious weakness of the PA, remember, is that it merely stipulates that pleasures (*qua* teleological subordinates) cannot be bearers of final value; it gives no reason for supposing this to be true. But if my interpretation of the passage as a whole is correct, the argument that directly follows the PA provides a missing reason: teleological subordinates cannot be final goods because, if they were, the value of the ends at which they aim would be effectively obliterated. So my interpretation not only attributes to Socrates an interesting pair of arguments; it also provides a way to see how the second argument supports the first.

3.5. Some Exegetical Objections and Replies

So far I have tried to argue that both the PA and the argument that supplements it are designed to show that pleasures, insofar as they are teleological processes, cannot be final goods. But in the first chapter I argued that, according to Socrates, *all* pleasures are teleological processes. From these two claims it follows that no pleasure is a final good. Most commentators resist attributing this conclusion to Plato, however, and their opposition is motivated by two independent considerations. First is the thought that Plato never intends to apply EM to all pleasures. On this view, Plato recognizes that EM cannot account for at least some pleasures—such as those of anticipation or virtue—and so implicitly leaves these pleasures untouched by the PA. Second is the thought that Plato, at other places in the *Philebus*, indicates that he wants to include at least some pleasures in the set of final goods. On this view, Plato cannot apply the PA's conclusion universally without contradicting positions he adopts elsewhere in the dialogue. I have already argued at length (in chapter 1) that the first worry is unfounded. The second worry is more serious, I believe, and so I devote the rest of this chapter to defusing it. But before moving on to that task, I think it will be worthwhile to deal briefly with some reasons that have been given for denying that Plato himself even endorses the PA.

Many scholars have correctly pointed out that Socrates does not take credit for the PA's crucial first premise (that is, the *process theory* = PT). Instead he attributes it to

an unnamed group of "elegant minds" (κομψοί: 53c4-7; cf. 54d4-e2). Some have taken this as evidence for the view that Plato, by distancing Socrates from the theory in this way, is effectively signaling that he does not entirely subscribe to it.[109] On this view, Socrates implicitly treats the "elegant minds" as he treats the "clever scientists" (δεινοὶ περὶ φύσιν) at 44b9—that is, as merely temporary allies in a common struggle against hedonism. But this cannot be right. Though it is true that Socrates never explicitly endorses the view of the "elegant minds," he takes great pains to warn Protarchus *not* to believe the "clever scientists" when they claim that every pleasure is only an escape from pain (44c1-5). Moreover, he explicitly accuses the "clever scientists" of analyzing pleasure "in an artless fashion" (οὐ τέχνῃ, 44c6) and then suggests quite clearly both that and why his own view parts company with theirs (44d2-5). Socrates does nothing like this with the "elegant minds." He voices no disagreement with their theory and twice expresses strong gratitude to them for publishing it (53c7; cf. 54d6).[110] That Socrates does not claim to be the theory's author is itself neither here nor there. For Plato frequently distances Socrates from the source of doctrines that Socrates—and Plato himself—clearly endorse.[111] This simply does not qualify as sound evidence for the claim that Plato is suspicious of PT.

Others claim to find evidence for this in the text of the PA itself. They maintain that Socrates, in his exposition of the PA, uses language which implies that he considers

[109] Hackforth (1972, 105-6) and Carone (2000, 265).
[110] *Contra* Carone (2000, 265).
[111] In the *Philebus* Socrates refuses to take credit for the "Promethean method" of collection and division (16c1-6) and the "dream solution" to the conflict between pleasure and intelligence (20b6-7). Surely this

PT to be a mere hypothesis: he twice refers to PT with the conditional phrase "*if in fact* pleasure is a process of becoming" (ἡδονή γε εἴπερ γένεσίς ἐστιν, 54c6, 54d1). According to Hackforth (1972, 105-6), the use of εἴπερ here is a sign that Plato—because he is skeptical of PT—insists only on the *validity* of the PA, and not on its soundness.[112] But this reads far too much into the phrasing. The use of εἴπερ in the context of a deductive argument can often function as a mere reminder of the argument's logical structure. Socrates might just as well be emphasizing the role of PT in the PA's crucial inferences. (See Aristotle's *De Caelo* 1.24 for a good example of this.) The language with which Socrates advances the PA yields no independent reason to suspect that Plato is reserving his judgment about its quality.

Many commentators—Hackforth (1972, 106-7) foremost among them—think that there are other textual grounds, elsewhere in the *Philebus*, for supposing that Plato does not fully endorse PT. In their view, the conclusion of the PA conflicts intolerably with several other of Plato's explicitly stated positions in the dialogue. Let us examine these apparent conflicts in order to see whether they can be resolved without casting suspicion on Plato's endorsement of the PA. I shall discuss them below, one by one, in order of difficulty.

(1) At 13b1 Socrates claims that some pleasures are good, which seems to suggest that he would deny the conclusion of the PA.

---

should not be taken as a sign that Plato does not endorse either one. For similar examples of deferred authority in other dialogues, see *Meno* (81a-b), *Symposium* (201d), and *Phaedrus* (244a).

[112] Cf. Carone (2000, 265).

This is a superficial worry. Though Socrates does call some pleasures good here, he makes the claim without fanfare and the conversation at this early stage is largely pre-theoretical. Since Socrates has yet to introduce any of the distinctions that fuel his analysis in the latter part of the dialogue, the predicate "good" in this context is not yet loaded with the same metaphysical weight it bears in the PA. By this remark Socrates might just mean that some pleasures are beneficial, where this implies nothing more than that they have a net positive instrumental value. And this is entirely consistent with both PT and the conclusion of the PA.

(2) Early in the dialogue Socrates develops the idea that some pleasures are "different from" and "opposed to" others (12c1-14a9). This might be thought to be inconsistent with PT, which treats pleasure as an undifferentiated unity.

This rests on a confusion that I tried to expose as such in chapter 1. That some pleasures are "different than" or "opposed to" others does not imply that there is no generic property that they all share *qua* pleasures. As Tuozzo (1996, 495-6) notes, Socrates explicitly compares pleasure-difference to shape-difference (12e3-13a5). Yet as early as the *Meno* Socrates recognizes that different shapes may be "opposed to" one another and yet share a single real definition.[113] And, as I have already argued, the PA draws general conclusions about pleasure based on the properties that all pleasures share in virtue of their real definition. So there is no reason to think that what Socrates says about the differences between pleasures conflicts in any way with PT as it stands.

[113] See above, section 1.2.

(3) In the final stages of the dialogue, Socrates explicitly includes some pleasures among the ingredients of the good life (e.g., 66c4-6), and this might be thought to imply that the included pleasures are final goods.[114]

Though it is clear that Plato wants to bring some pleasures into the good life, it is not entirely clear *why*. If he thinks that only final goods can be included in the good life, then of course his claim that the good life will have pleasure in it conflicts with the conclusion of the PA. But if he thinks that the inclusion criteria for the good life are lax enough to permit instrumental or otherwise derivative goods as well, then he need not hold that the pleasures to be included in that life are final goods. He might hold that they should be included, if at all, either because one cannot acquire final goods without them, or because one cannot live a recognizably human life without them.

Evidence that Plato in fact holds this latter view emerges most clearly when Socrates first takes up the task of constructing the good human life. There he claims that if some pleasures turn out to be "necessary" (ἀναγκαῖαι) then they should be included in the good life straightaway (62e3-10).[115] It is not immediately clear what Socrates means by this, but a bit earlier he suggests that one reason to include any item whatsoever, whether it be a kind of knowledge or a kind of pleasure, is that a human being cannot live a human life without it. After a protracted debate, he and Protarchus ultimately agree that the various forms of practical knowledge should be included just because each of them is "necessary [ἀναγκαῖον] at least if our life is going to be any kind of life at all" [εἴπερ γε

[114] This worry is raised most forcefully by Carone (2000, 269).
[115] Cf. *Republic* 558d ff., where Plato suggests that some pleasures necessarily come along with the satisfaction of basic biological needs.

ἡμῶν ὁ βίος ἔσται καὶ ὁπωσοῦν ποτε βίος] (62c3-4, cf. 62b8-9).[116] So the suggestion here seems to be that a human life is effectively unlivable without at least some practical knowledge and without at least some pleasure.[117] A bit later, in an odd exchange at 63b2-64a5, Socrates asks the various types of intelligence—which have been personified for rhetorical effect—whether they would be willing to share space with any pleasures in a life. They reply that the only types of pleasure they would be willing to live with are those that either go along with or promote states that are otherwise thought to be valuable on their own. These genuinely valuable states include health (63e4), virtue (63e5), and—as Socrates makes clear at the very end of the dialogue—knowledge itself (66c4-6).

Taken together, these passages strongly suggest that Plato includes pleasures in the good human life either because they are necessary components of *any* human life, or because they are somehow attached to final goods. And if these are Plato's only grounds for bringing pleasures on board, then he can endorse the conclusion of the PA without being guilty of self-contradiction.

(4) At 52c1-d1 Socrates asserts that some pleasures belong to the class of measured things. Since he also thinks that the class of measured things is coextensive with the class of final goods, he seems committed to the claim that these pleasures, at least, are final goods.

Though it is never entirely clear what is supposed to belong to the class of measured things, it is quite clear that, for Plato, whatever belongs to that class is a final good. So if he is implying in the contested passage that some pleasures belong to the

[116] As Hackforth (1972, 127) puts it, "this decision is taken out of regard for the needs of practical life."

measured class, then he seems committed to the claim that these pleasures—whichever they may be—are also final goods. The passage itself runs as follows in the text:

> Since we have now properly separated the pure pleasures and those that can certainly and rightly be called impure, let's add, in our account, unmeasuredness to the intense pleasures [ταῖς μὲν σφοδραῖς ἡδοναῖς ἀμετρίαν], and measuredness to those [pleasures] that are not [intense]. And let's set down that the pleasures that receive the large and the intense [τὰς τὸ μέγα καὶ τὸ σφοδρὸν δεχομένας] … belong to the class of the unlimited, the more and less, which goes through both body and soul. [And let's also set down that] the pleasures that do not [receive the large and intense] belong to the class of measured things [τῶν ἐμμέτρων]. (52c1-d1)

As Cooper (1999, 150-153) and Carone (2000, 267-270) have pointed out, this remark seems to imply that the pure pleasures belong to the class of things specified in the fourfold division as combinations of limit and unlimited. But as I argued in chapter 1, all of the items in this so-called "mixed" class are final goods. So if this passage is to be trusted, Socrates must think that the pure pleasures are final goods.

There is reason to doubt that this is Plato's considered view, however. First, it is independently difficult to make sense of the claim that, for Plato, there are some pleasures that belong to the mixed class. According to Plato's EM, every pleasure is a process in which a psychological equilibrium state is restored. But EM also holds that the equilibrium states being restored in each case are *themselves* members of the mixed class. Indeed, when Socrates sets out his fourfold division, he implies that *only* equilibrium states belong to the mixed class. So if the pure pleasures too are restorations of equilibria—as I argued in chapter 1—then evidently they cannot *themselves* be equilibria.

[117] Cf. Frede (1993, lxvi) and Cooper (1999, 159-160).

It is of course possible to read this passage as evidence against the view that EM is designed to account for all types of pleasure. But as I have already argued, the weight of the evidence is much stronger in favor of that view. Moreover, Socrates at one point *explicitly denies* that pleasure can belong to the measured class. "Let us remember [μεμνώμεθα]," he says, "that … pleasure itself is unlimited, and belongs to the genus that in and by itself neither has nor will ever have beginning, middle, or end" (ἡδονὴ δὲ ἄπειρός τε αὐτὴ καὶ τοῦ μήτε ἀρχὴν μήτε μέσα μήτε τέλος ἐν αὑτῷ ἀφ' ἑαυτοῦ ἔχοντος μηδὲ ἔξοντός ποτε γένους, 31a7-10). This claim cannot easily be reconciled with the idea that *only some* pleasures belong to the "class of the unlimited."

So we have an apparent conflict. Perhaps the best way to proceed at this point would be to leave aside the question of whether pleasures can be final goods, and examine some of Socrates' cues about how and why a pleasure might be worth pursuing in any case. During his discussion of the different types of pleasure and pain, Socrates suggests, almost in passing, that there are some pleasures it might be worthwhile to have even if they are not themselves goods. He hopes that, as his conversation with Protarchus proceeds,

> it will become clear about pleasure whether the whole class is to be welcomed [ἀσπαστόν], or, if this title should be given to one of the other classes we have mentioned, whether pleasure and pain, just as with hot and cold and the rest, are to be welcomed at one time and not at another, since, although they are not good, they sometimes receive the nature of the goods [ὡς ἀγαθὰ μὲν οὐκ ὄντα, ἐνίοτε δὲ καὶ ἔνια δεχόμενα τὴν τῶν ἀγαθῶν ἔστιν ὅτε φύσιν]. (32c8-d6)

The suggestion here seems to be that pleasures, though not themselves goods, might be worth pursuing if they somehow "receive the nature" of goods. Socrates does not elaborate, but it is fairly easy to see what he is getting at. By drawing a distinction between things that are "to be welcomed" and things that are "good," he hopes to fend off an inference from the claim that no pleasures are goods to the claim that no pleasures are ever worth having. In his view, I take it, some pleasures are bound up with final goods in such a way that they are worth having, even if they are never worth having *in themselves*. If this is right, then Socrates is here preparing Protarchus for the idea—to be developed shortly—that it is only in virtue of a pleasure's connection with genuine non-hedonic goods that it is fit to be had at all. Though Socrates remains silent at this stage about the precise nature of this connection, he suggests elsewhere that pleasure, which is itself "unlimited," must be "tied down by a limit" (ὑπὸ τοῦ πέρατος δεδεμένων) in order to become beneficial (27d7-10).[118] The thought here seems to be that pleasure, though itself unlimited—and hence not a candidate for final value—can nonetheless become beneficial by being limited in some way.[119]

All of this suggests, then, that for Plato a pleasure can be worth pursuing even if no pleasure "itself by itself" can be a genuine good. Pleasures "to be welcomed" are worth pursuing at all only insofar as they are connected in some as-yet-unspecified sense with

[118] When Philebus appeals to the goddess Aphrodite to bolster his claim that pleasure is the good, Socrates counters that "it is the goddess herself who sees the hubris and wickedness in everything and, seeing that there is no limit in pleasures … imposes law and order upon them as a limit. You claim that she ruins them [ἀποκναῖσαι], but I, on the contrary, say that she saves them" [ἀποσῶσαι] (26b7-c1).

[119] This seems to be what Annas (1999, 152-153) is after when she claims that pleasure, for Plato, "is raw material which enters the [proper] combination in a way determined by reason. Reason dominates the

those things that are worth pursuing regardless. As I see it, there is an interesting way to cash out the relation that must obtain between a pleasure and a genuine good in order for that pleasure "to be welcomed." Drawing on my discussion in chapters 1 and 2, we might suppose that a pleasure can "receive the nature of the goods" if it is properly attuned to the equilibrium state at which it aims. And this, I take it, can be understood in two different ways. First, a properly attuned pleasure can be taken to be a psychological restoration that does not generate addictive conditions in the agent undergoing it. If the pleasure is in this sense moderate, and not intense, then it allows the agent to arrive at his relevant equilibrium state without setting off an endless chain of pleasures mixed with pain. Second, and more interestingly, a properly attuned pleasure can be understood as one that has the right sort of content. When a pleasure-attitude is appropriately aimed, it takes as its intentional object a genuine bearer of final value, and thereby "receives the nature of the good" at which it aims.[120] In either case, the pleasure "to be welcomed" has achieved this status in virtue of its connection to something that, for Plato, has genuine final value. But in neither case does the *pleasure itself* have genuine final value. To make this further move would merely invite confusion.

---

mixture in that it determines what the limit is that is to be imposed on pleasure for pleasure to be able to contribute to the happy life." Cf. Frede (1992, 454).

[120] This interpretation is broadly compatible with Cooper's (1999, 161-163). According to Cooper, "whether [a physical pleasure] is intense will depend upon what in the experience is being enjoyed. If what is enjoyed is not, or not merely, the release from tension, then the pleasure […] taking place then will not belong to [the unlimited] class" (162). So long as the object of one's enjoyment when drinking is "the activity of maintaining [one's] physical substance and preserving [one's] bodily health," then one's pleasure is properly limited. Thus Cooper and I are in agreement to this point. Pleasures can qualify as "to be welcomed" in virtue of having the right sort of intentional content. But as we have seen, Cooper goes further, suggesting that these sorts of pleasures are to be understood as choiceworthy for their own sake. And here he and I part company.

I believe this is the best general way to save Plato's account from what would otherwise be a hopeless state. In the problematic passage above, Socrates should not be interpreted as saying that some pleasures belong to the mixed class in the same way that *health*, say, belongs to the mixed class. He should rather be interpreted as claiming that some pleasures can "receive" measure by being appropriately connected to things in the mixed class, and so thereby become choiceworthy—though not, to be sure, choiceworthy for their own sake. This preserves the coherence of Plato's view without detracting from what I take to be his main point.

(5) At 21d6-22b8 Socrates delivers an argument based on a hypothetical rational choice between different possible lives, and this argument seems to rest on the idea that some pleasures are final goods.

Whereas the first four conflicts proved themselves amenable to resolution in fairly short order, this final conflict is so nettlesome that I will need the rest of the chapter to untangle it.

3.6. The Choice Argument: An Introduction

One of the most celebrated arguments from Plato's *Philebus* ends with the conclusion that the best human life is neither wholly without intelligence nor wholly without pleasure, but includes at least some types of each (20b6-22b8; 60b7-61a2; and 66e7-67a12). I call this the *choice argument*, since it hinges on the results of a

hypothetical choice among three different ways of life. According to Plato, any suitably situated agent presented with this choice will reject both the completely unintellectual life of pleasure and the completely pleasureless life of the mind, provided that a third "mixed" life is available—one in which pleasure and intelligence each find a prominent and comfortable place. Since Plato takes the result of this hypothetical choice to show that the best human life is indeed a mixed one, he must think that the agent is fully justified in rejecting the first two alternatives and accepting the last. But it is not entirely clear what the agent's justification is supposed to be, let alone why Plato gives the choice his seal of approval. On what grounds does the agent decide to forego the two pure lives in favor of the mixed one? And why does Plato endorse the agent's decision?

Most commentators assume that Plato approves of this choice because he holds that at least some types of both pleasure and intelligence are final goods. Thus the first two lives are to be rejected precisely because each of them is missing a psychological feature that is worth pursuing for its own sake. John Cooper (1999, 152) takes this line in a recent collection of essays, arguing that "both the pleasure and the knowledge [in the mixed life] are recognized to be good things in themselves (things without which life would not be satisfactory for a human being, not merely because of the further things their absence would make unavailable, but because of the very natures of the pleasure and the knowledge themselves)." Terence Irwin (1995, 335) agrees: "if some forms of rational activity necessarily involve pleasure, then the life that includes them involves pleasure. Socrates wants us to set this possibility aside, however, and focus on the question of

whether the pleasure is itself part of what is good about a good life."[121] Cooper and Irwin are advocates of what I have called the *liberal view*, according to which Plato allows at least some pleasures into the class of final goods.[122]

In what follows I will argue that the liberal interpretation of the choice argument (CA) is mistaken. Though the CA does establish that certain types of pleasure and intelligence are to be included in the best human life, the argument itself cannot explain *why*. The best it can do is establish that the relevant types of pleasure and intelligence must be included *either* because they are final goods *or* because they are instrumental goods that one must acquire if one is to acquire final goods. If this reading is correct, then the CA by itself does not commit Plato to the view that some types of pleasure are bearers of final value. Indeed, as I have already argued, this is a view that Plato explicitly rejects in the *Philebus* (54c1-d7). Since Plato's anti-hedonism is incompatible with the claim that some pleasures are final goods, he must hold that the types of pleasure to be included in the good life are there only because they are somehow necessary for it. So while Plato does insist that the pleasureless, painless life of the intellect is not the best human life, he does so only because he believes that we as human persons are incapable of living it. Or so I shall argue.

Central to the general framework of the CA is the idea that a sufficiently discriminating agent will prefer one among a certain range of items, provided that this

---

[121] According to Irwin (1995, 336), Plato's claim is that "pleasure is another part of happiness in its own right."

[122] On this point, at least, Carone (2000, 258-261) follows Cooper and Irwin. For an independent endorsement both of the liberal reading of the argument and of the argument so read, see Lemos (1994, 53).

agent deliberates properly according to specific criteria of choice. This formulation leaves a number of questions open, however, and our first task as interpreters is to figure out how Plato wants us to answer them. What sort of agent is making the choice, and under what circumstances? What is the agent's range of alternatives? And what criteria are to govern the agent's choice? Plato sketches out answers to all three of these questions at various preliminary stages of the argument, and each answer both demands and repays careful attention.

3.7. The Criteria

The hypothetical choice that underwrites the conclusion of Plato's argument is not an arbitrary one. It is governed by a set of interrelated criteria expressing standards which the chosen object, if chosen appropriately, must satisfy. These criteria effectively provide the membership conditions for what Socrates calls the "class of the good" (τὴν τἀγαθοῦ μοῖραν, 20d1). According to Socrates, this class has three essential properties. It must be (1) complete (τέλεον, 20d1-2); (2) sufficient (ἱκανόν, 20d4); and (3) such that

> anything that knows it aims at and pursues it (πᾶν τὸ γιγνῶσκον αὐτὸ θηρεύει καὶ ἐφίεται), and wants to take and possess it for its very own, and cares for nothing else except the things acquired along with goods (καὶ τῶν ἄλλων οὐδὲν φροντίζει πλὴν τῶν ἀποτελουμένων ἅμα ἀγαθοῖς) (20d7-10).[123]

[123] On the translation of this passage, see White (2002, 185n87).

Taken as a whole, these three membership conditions strongly suggest that the "class of the good" (CG) contains all and only items that are worth pursuing for their own sake. The first two criteria imply that CG neither lacks any of the goods there are, nor includes anything that is not a good.[124] The third criterion specifies more precisely what sort of value these goods are supposed to have. It sketches a crucial distinction between the things we care about as soon as we recognize them, and the things we care about only because, only if, or only when we get them "along with" the things we care about anyway. This distinction suggests not just that CG includes all and only the things we are justified in pursuing for their own sake, but also—and perhaps more importantly—that we are justified in pursuing things that do not belong to CG. There are other things worth caring about, even though their value is somehow parasitic on the value of the members of CG. Because Plato recognizes two different classes of choiceworthy objects, and because only one of these classes is composed of final goods, it is crucially important for us to determine more precisely what these classes are and how they are related. Otherwise we run the risk of mistaking derivative goods for final goods whenever the properly qualified agent chooses to have both. So what is this second class of valuable things, and how exactly is it related to the first?

Christopher Bobonich (1995, 122) suggests that the second class is composed of what he calls "dependent goods," where dependent goods are defined as things that are "good for [their] possessor only if she possesses something else" (i.e., one or more of the

[124] On the first two criteria, see Hackforth (1972, 32), Davidson (1990, 139-147), and Irwin (1995, 332-333).

primary goods).[125] Bobonich finds support for this reading in the first book of the *Laws* (631b-632a), where Plato draws an apparently similar distinction between "human goods" (e.g., health, beauty, strength, and wealth) and "divine goods" (i.e., the virtues). Plato's examples in that text suggest that he holds what Bobonich calls "the dependency thesis," which entails that the so-called "human goods" are genuinely good for an agent only if that agent has already acquired the "divine goods." Put simply, the human goods depend on divine goods for their very goodness. According to Bobonich, then, Plato's thought is that human goods such as health and strength are easily enlisted in the service of vice, and so are genuinely worth pursuing for an agent only if that agent already has the divine good of virtue. Virtue, on the other hand, is worth pursuing in any case and for its own sake. If Bobonich is right, then this is the same distinction that informs the third criterion in Plato's choice argument, and it fixes the relation between the two types of value at issue in that text.

Though Bobonich's interpretation of the *Laws* passage is interesting and plausible, it seems forced as a reading of the third criterion in the *Philebus*. If the second value class contains things that are not themselves goods, but that we acquire "along with" goods, then it is hard to see how they could *become* good for us when (and only when) we have acquired the members of CG. It seems more appropriate to assume, given the language of this text, that the members of the second value class are certain non-goods that we acquire (for whatever reason) when we acquire goods. So instead of calling these items "dependent goods," I propose to call them *complements*. If this alternative reading

[125] See also Bobonich (2002, 123-159, esp. 153-159).

is correct, then the acquisition of goods is somehow tied up with the acquisition of complements, and this is supposed to explain why complements are appropriate objects of care and pursuit.

But this gloss on the relation between goods and complements is still quite vague. If the relation could be clarified by some illustrative examples from the *Philebus*, it would have a better claim on our attention. Consider, then, the many occasions in the dialogue when Socrates claims that certain types of cognitive activity are bound up with certain types of pleasure (at least in human beings). He claims that there is pleasure in:

(1) anticipating that one will satisfy an appetite (32b9-c2);

(2) remembering having satisfied an appetite (33c5-6);

(3) believing that some evil has befallen one's enemies (49d3-4);

(4) learning (51e7-52a3);

(5) exercising abstract knowledge (63e3-4); and

(6) applying knowledge to practical tasks (66c4-6 and 56b4-c2 with 51c1-6).

The connection between pleasure and cognitive activity in these cases is clearly not accidental. In each case, Socrates thinks, a specific type of rational consciousness is necessarily tied to a specific type of pleasure. There is no learning without the pleasure of learning, and *vice versa*. What this suggests, then, is that if some types of cognition are final goods, then the types of pleasure correlated with them are *complements* of these

goods. Likewise, if some types of pleasure are final goods, then the types of cognition correlated with them are their complements. For in each case, if you acquire one, you acquire the other "along with" it. So Plato's own treatment of the relation between certain pleasures and certain cognitive activities illustrates, in a fairly straightforward way, the distinction between complements and goods—without committing to any claim about which is which.

This complementarity thesis (CT) differs from Bobonich's dependency thesis (DT) in subtle but important ways. According to DT, remember, the choiceworthiness of things in the subsidiary class depends, in each case, on the agent's having already acquired genuine goods. But according to CT, the choiceworthiness of things in the subsidiary class depends on their *having to be acquired along with* primary goods. Since we necessarily get complements along with genuine goods, complements are part of a package deal, as it were. If we get rid of them, we also get rid of the goods. So a hypothetical life that has all the goods but lacks all the complements belonging to these goods is an *impossible* life—at least for human persons. While both CT and DT imply that no items in the subsidiary class are worthy of ultimate pursuit, each thesis posits a different reason for pursuing them. According to DT, the agent has a reason to care about items in the secondary class only if the agent has already acquired genuine goods; but on CT, the agent *always* has reason to care about items in the secondary class—not for their own sake, but for the sake of the genuine goods they come along with. Since the agent has a reason to pursue these things even though they are not genuine goods, we must be careful not to assume that

whatever features belong to the life chosen by the perfectly rational agent must themselves belong to the class of final goods. For this life will include both final goods and their complements.

3.8. The Alternatives

After setting out the criteria that the hypothetical chooser must adopt, Socrates establishes the range of things available to be chosen. Here he is not admirably clear. The most pressing question for the interpreter is whether the chooser is to pick among potential *lives* or among potential *features* of lives. At the beginning of the dialogue, Socrates seems to recognize this distinction. "Each of us," he says, referring to himself and the hedonist Protarchus, "will try to show that some state and condition of soul [ἕξιν ψυχῆς καὶ διάθεσιν] is able to provide the good life [τὸν βίον εὐδαίμονα] for every human being" (11d4-6). Here Socrates plainly grasps the difference between a life and one of its psychological features. After all, his claim is that the conflict between the hedonist and the intellectualist turns on the question of whether pleasure or intelligence is the psychological feature of a life that *makes* it good. Later, in the immediate run-up to the CA, Socrates implies that the results of a hypothetical choice between lives can reveal the value of the characteristic features of these lives. In his view, the CA will establish that "neither [pleasure nor intelligence] is the good, but rather some third thing, different from them and superior to both" (20b6-9). The suggestion, again, is that the CA will bear on

the value of these features, even though the hypothetical agent will evidently be choosing among lives, not features. In order for this to work, Socrates must assume that he can infer something about the value of life-features from the result of hypothetical life-choices. The closest he gets to articulating this assumption comes, once again, at a preliminary stage of the argument. Here he stipulates some rules about the composition of the lives to be considered by the hypothetical chooser:

> Let there be neither intelligence in the life of pleasure nor pleasure in the life of intelligence [μήτε ἐν τῷ τῆς ἡδονῆς ἐνέστω φρόνησις μήτ᾽ ἐν τῷ τῆς φρονήσεως ἡδονή]. For if in fact either of them [εἴπερ πότερον αὐτῶν] is the good, it must not need anything else in addition. But if one or the other should appear to be lacking [δεόμενον], then I suppose this can no longer be the real good for us [τὸ ὄντως ἡμῖν ἀγαθόν]. (20e4-21a2)

Even a careful reader of this passage might not notice that Socrates, in the second sentence, introduces an ambiguity into the discussion. It is not clear at this point whether he is suggesting that his test will reveal which of the two *features* is the "the real good for us" or which of the two *lives* is "the real good for us." But in the lines that follow this passage, Socrates makes it fairly explicit that what he takes himself to be testing directly in the argument are *lives*, not *features*.[126] His suggestion, then, seems to be that if either of the unmixed lives is lacking something, then it—that is, the *life*—is not "the real good for us." So Socrates has apparently shifted the focus of his question. Instead of asking whether one of the two features is "the good," he is asking whether one of the two unmixed lives is "the good [life] for us." Despite this shift in focus, Socrates apparently

thinks that the answer to the second question will bear directly on the answer to the first. The challenge now is to figure out *why* he thinks this.

One initially attractive way to reconstruct his reasoning is to think of a life as something that includes its features roughly as a set includes its members. From this standpoint we might test whether or not some set of features F is identical to the "class of the good" (CG) by constructing a life L that is coextensive with F, and then determining whether or not L meets the three membership conditions of CG (completeness, sufficiency, and choiceworthiness). Our goal, then, would be to discover whether or not F is coextensive with CG. So if we are trying to decide whether or not pleasure is the good—and if we assume that the only competitor for the title of the good is intelligence—then we can perform something like the following test. First, construct a life L/P that is coextensive with the class of pleasures. Then decide whether L/P is "in need of" some kind of intelligence in addition to the pleasures it already has. If so, then L/P is not coextensive with CG, since it is either insufficient or incomplete. Therefore pleasure is not the good. The same test can then be applied in a similar fashion to intelligence. If these tests show that neither pleasure nor intelligence is the good, then at least some types of both intelligence and pleasure must belong to CG.

Advocates of the liberal interpretation seem to think that this, roughly speaking, is how the CA unfolds. But this way of reconstructing the argument rests on a crucial assumption, namely, that the membership conditions for "the good life for us" are the

[126] See, for example, 21a8-9, 21d6-7, 22a9-b6. In each of these texts the form of the question is essentially the same: would you choose life L, or is L missing some feature that any life worth choosing must have?

same as the membership conditions for the "class of the good." In other words, it presupposes that if pleasure is the good, then a life of pleasure unmixed with intelligence must be the good life for us. Yet this way of reconstructing the argument evidently ignores Plato's distinction between goods and complements. If that distinction is to be taken at all seriously, then "the good life for us" will contain not only all the items that belong to the "class of the good," but also the *complements* of these items, if there are any. Given that the good-making features of a human life necessarily "come along with" other features which are not themselves good-makers, any life that fails to have these other features will not be choiceworthy for us. If this is right, then a life might be "in need of" something in order for it to be choiceworthy for us, yet not be "in need of" any final goods. It might rather be "in need of" complements. The liberal interpretation has it that if pleasure is the good, then the unmixed life of pleasure is the good life for us. But if we take the distinction between goods and complements seriously, then this is a bad inference. The correct inference is that if pleasure is the good, then a life of pleasure *mixed with pleasure's complements* is the good life for us. For if it lacked these complements, we would presumably be unable to live it.

The choice test, then, ought to be formulated differently. Instead of testing a life that is coextensive with the set of features in question (F), we should test a life that is coextensive with the union of F and the set of the complements of F. (I will represent this union as C<F). So we should construct a life L/C<F that is coextensive with the union of F and C<F, and then determine whether or not L/C<F meets the membership conditions

of the union of CG and the complements of CG (C<CG). But Socrates for some reason fails to do this. Instead he stipulates at the outset that each life must be tested in its unmixed form, without any potential complements added. He removes all intelligence from the life of pleasure and all pleasure from the life of intelligence. The problem with this is obvious: if the choice test is designed to reveal whether any types of pleasure or intelligence are final goods, then submitting only the unmixed lives to this test will in each case yield inconclusive results. For even if one of these unmixed lives fails the choice test, the reason for its failure is underdetermined. It might lack final goods, but then again it might just lack complements. So even if the unmixed life of pleasure were to fail the choice test, we could not infer from this that some types of intelligence are final goods.

Defenders of the liberal interpretation might object at this point that, for this reason alone, my reading cannot make sense of the argument as it appears in the text. Socrates clearly takes the CA to have established that "neither pleasure nor intelligence is the good." So if the actual argument fails to license any conclusion about whether some types of pleasure or intelligence are goods, then either Socrates is hopelessly confused about what his own argument can accomplish or my reading is incorrect. Or so the objection goes. But this line of reasoning depends on a dubious assumption about what Socrates means when he claims that "neither pleasure nor intelligence is the good." According to the liberal interpretation, he means that neither pleasure nor intelligence—understood now as sets of psychological features—is coextensive with the "class of the good." If this is what he means, then my interpretation does not jibe with

Socrates' own interpretation. But this need not be what Socrates means. When he declares that "neither pleasure nor intelligence is the good," he might rather mean that neither of the two sets of features exhausts the features of a good *life*. In other words, he might think that if "pleasure is the good," then a life of pleasure unmixed with intelligence is the good life; no further assumptions about the final value of these features need be built into the claim. And if this is what he means, then the CA does indeed establish that neither pleasure nor intelligence is the good, since each of the two unmixed lives is missing something necessary to make it choiceworthy—whether that be a good or a complement.

In the end, the conflict between the liberal interpretation and mine will be settled only by a close analysis of the argument itself. But before I turn to the text, I think it would be worth considering for a moment just how unreasonable Socrates would have to be if the liberal interpretation were correct about what he expects the CA to accomplish. According to the liberals, remember, Socrates must believe that the good-making features of a life—whether they be types of pleasure or types of intelligence—can succeed in making that life good even in the absence of any other features. This is unreasonable, I think, because it rules out in advance the possibility that a good human life has features which are not themselves good-makers, but which any human life must have in order to be a recognizably *human* life.[127] In my discussion of the goods/complements distinction

[127] Lemos (1994, 53) canvasses, and then rejects, a slightly different line of resistance: "One objection to [the argument as interpreted by the liberals] is that the reason no one would find the change into an immortal cricket or a very long-lived imbecile desirable is that what exists afterward "would not be me." In other words, such a change would be destructive of the person." This objection has no force, according to Lemos, since there is "no good reason to think it logically impossible for some misfortune, some accident or disease, to befall me that would result in my being such a contented imbecile." On the assumption that such an "imbecile" has the same level of cognitive ability as a "cricket," I find it very odd indeed that

above, I pointed out that Socrates himself seems aware that pleasure and cognition are, in human beings at least, necessarily intertwined. If this is his view, then he should not make the claim attributed to him by the liberal interpretation.

To see more clearly why this is so, assume for the sake of argument that the various types of cognition are the only final goods there are. By a further hypothesis that Socrates himself accepts, human beings cannot experience at least some of these types of cognition without also being pleased. But then a life made up of all the various types of cognition without any types of pleasure would be an impossible life for humans to live—and hence, on any reasonable account, would not be choiceworthy for us.[128] So Socrates should be able to see that such a life, despite its not being choiceworthy for us, has the same good-making features as the life that *is* choiceworthy for us. According to the liberal interpretation, Socrates does not see this possibility. On my interpretation, he does—and I take this to be a point in its favor.

3.9. The Argument

Let us now turn to the argument itself. The CA is divided into two parts which, taken together, are supposed to establish that neither the unmixed life of pleasure nor the

---

Lemos thinks he as a human person could survive such a transition—though of course I agree that the issue would not be settled by appeal to what is "logically possible." Frede (1993, xxxii), I think, is more on the mark: "if there is to be anything worth calling a human life, there has to be memory of past pleasures and full comprehension of present and possible future ones."

[128] I am not assuming here that *any* life that is impossible for us to live is not choiceworthy for us. As Terence Irwin has pointed out to me in correspondence, our inability to live a life completely free of illness does not seem to make such a life any less choiceworthy for us. The issue, rather, is whether or not the life

unmixed life of cognition is choiceworthy for us. Socrates begins the first part by asking Protarchus whether he would choose "to live his entire life enjoying the greatest pleasures" [ζῆν τὸν βίον ἅπαντα ἡδόμενος ἡδονὰς τὰς μεγίστας] or whether he would "need something else in addition" [ἔτι τινὸς ἄν ... προσδεῖν] (21a8-12). When Protarchus denies that he would need anything else, the following exchange takes place:

> Socrates: But look, wouldn't you need some sort of intelligence or reason or calculation or anything else related to them?
> Protarchus: What for? In having pleasure I would have everything.
> Socrates: But in living this way always throughout your life, would you enjoy the greatest pleasures? [οὐκοῦν οὕτω ζῶν ἀεὶ μὲν διὰ βίου ταῖς μεγίσταις ἡδοναῖς χαίροις ἄν]
> Protarchus: Why not?
> Socrates: Given that you do not possess mind [νοῦν δέ] or memory or knowledge or true opinion, it certainly follows that you are ignorant, first, of this itself, whether you are enjoying or not, since you are empty of all intelligence. (21a14-b9)

Socrates goes on to add four further disabilities: Protarchus would be unable to remember any of his pleasures (21c1-2); to experience a single pleasure across time (21c2-4); to judge that he is pleased even when he is (21c4-5); and to calculate that he will have pleasure in the future (21c5-6). And from all this Socrates draws a conclusion that leaves Protarchus somewhat startled:

> Socrates: You would live not a human life [ζῆν δὲ οὐκ ἀνθρώπου βίον], but that of a mollusk or one of those other shelled creatures that live in the sea. Is this right, or can we think of any other consequences besides these?
> Protarchus: How could we?
> Socrates: But is such a life choiceworthy for us? [αἱρετὸς ἡμῖν βίος ὁ τοιοῦτος]

---

in question is impossible for us to live because living it would involve losing our status as human persons.

> Protarchus: Socrates, this argument has left me completely speechless at the moment. (21c6-d5)

According to the liberal interpretation, the purpose of all this is to show Protarchus that cognitive activity has some independent final value, and to convince him that the unmixed life of pleasure is deficient as a result.[129] But this reading is underdetermined by the text of the argument. At the crucial point of exchange, Socrates asks Protarchus whether he would lead the life of greatest pleasure if he were bereft of cognition. And this question can be read in two different ways. On the liberal interpretation, Socrates is asking whether Protarchus would *choose* to have such non-cognitive pleasures, given that he *could* have such pleasures if he so chose. But χαίροις ἄν can also be read as a potential optative, and on this reading Socrates is not asking whether Protarchus would *choose* to have such pleasures; he is asking whether Protarchus himself *could* have such pleasures, even if he chose to. On this alternative reading, Socrates is skeptical that Protarchus *qua* human person *could* live a crustacean life of maximum pleasure.

As far as I know, the only commentator who attempts to develop something like this alternative reading is Gosling (1975, 183). Gosling's view, however, is that Socrates is trying to show Protarchus that he would not be able to live the life of greatest pleasure if he were wholly unintelligent. For Gosling, then, the point of the argument is that shellfish cannot experience the full range of pleasures that are open to humans, and so

[129] This is how Moore (1993, 139-141) reads it as well, claiming that Plato's point here is to show that pleasure is valueless for an agent unless the agent is at least conscious of being pleased.

cannot live a genuinely pleasure-rich life. Irwin (1995, 334) is right to suspect that there is something wrong with this interpretation. "Socrates," he writes, "would be dealing unfairly with the hedonist position if he meant to recommend these forms of rational consciousness simply as a means to greater pleasures. For he has already conceded that the unmixed life of pleasure includes the greatest pleasures." I agree that this is a problem for Gosling's view. But I do not believe it is a problem for mine. On my interpretation, Socrates is not suggesting that Protarchus needs intelligence in order to lead a life of greatest pleasure; he is suggesting that Protarchus needs cognition in order to lead a *human* life of greatest pleasure. The point of the argument is to show Protarchus that he, *qua* human person, could not lead a wholly unintellectual life of greatest pleasure, because a life of this sort is not sufficiently human. And if this is right, then Socrates need not alter his earlier claim that such a life has the greatest pleasures. He need only claim that the pleasures this life has—even if they are the greatest—involve no cognition whatsoever.

On my reading, then, the argument does not show that a *human* life of greatest pleasure—which would of course have to include various cognitive capacities—fails to be choiceworthy for us. Nor does it show that the unmixed life of greatest pleasure fails to be choiceworthy for a mollusk. It shows only that such a life is not choiceworthy *for a human person*. Does the failure of this life to be choiceworthy for us imply that the intellectual component missing from its pleasures is something good in itself? It does

not.[130] A hedonist should be entitled to include cognition in his ideal life, provided that he values it only as a complement of pleasure—as something that necessarily "comes along with" pleasure but is not good in itself. A similarly limited result emerges from the second part of the CA, which concludes that the pleasureless life of the mind is no more choiceworthy for us than the mindless life of pleasure. While the first part of the CA fails to show that any type of cognition is good in itself, the second part fails to show that any type of *pleasure* is good in itself.

The second part of the argument is more compressed than the first, presumably because Protarchus is already disposed to accept the idea that some sorts of pleasure belong in a good human life. Socrates asks him whether the life of cognition is complete, sufficient, and choiceworthy for all even if "it shares neither in pleasure … nor in pain, but is entirely insensitive [τὸ παρὰπαν ἀπαθὴς] of all such things" (21d9-e2).[131] Not surprisingly, Protarchus rejects the apathetic life as well. But once again, it is not obvious that Protarchus' genuine warrant for rejecting this life is what he thinks it is. If he thinks that the apathetic life fails the choice test because some types of pleasure are final goods, then—on my interpretation, anyway—he is mistaken. The choice test shows only *that* a life falls short of being choiceworthy, not *why*. So it is still possible, even after the CA is complete, for Socrates to maintain that certain types of pleasure are to be included in the

[130] Cf. Bobonich (1995, 121-122) and (2002, 157-159). *Pace* Irwin (1995, 333-335).
[131] It is worth noting that Socrates has made the argument even more forgiving to the exclusive life of intelligence by claiming that, for the sake of this competition, it should be understood as not only joyless but also painless. Although it is clear that the thoughtful life should lack pleasure in this contest, it is not at all clear why it should lack pain as well—after all, a hypothetical life full of intelligence might well be both pleasureless and painful. But none of this is relevant to the results of this particular choice test, since being a pained intellctual will be no more choiceworthy than being an apathetic intellectual.

good human life because they are complements of certain types of rational activity, not because they are good in themselves. Thus, contrary to the liberal interpretation, the CA leaves wide open the question of whether there are any genuinely valuable pleasures. This is a good thing from Plato's standpoint, I take it, since the CA (so interpreted) generates no conflict with the anti-hedonist conclusion of the PA. Plato can hold both that the best human life includes some pleasures and that no pleasures have final value. My reading, unlike the liberal one, makes the conclusions of the CA and the PA consistent.

3.10. Objections, Complications, and Refinements

So far I have tried to suggest that my interpretation of the CA is superior to the liberal one, if only because it makes Plato's position consistent across the dialogue. But as we shall see, matters are more complicated than this. When it comes time for Socrates to sum up what he takes the CA to have accomplished, he says a number of things that cause serious problems for both interpretations. Just after concluding that the pleasureless life of the mind is no more choiceworthy for us than the mindless life of pleasure, Socrates tries to summarize the argument's lesson:

> Socrates: So do we now understand the upshot of the present argument?
> Protarchus: Certainly. It is that, of the three lives on offer, two are neither sufficient nor choiceworthy for any human or animal [οὐδέτερος ἱκανὸς οὐδὲ αἱρετὸς οὔτε ἀνθρώπων οὔτε ζώων οὐδενί].
> Socrates: Well then, isn't it clear that neither of these two is the good? For otherwise it would be sufficient and complete and choiceworthy for all the plants and animals that are able to live in this way consistently throughout

their lives [ἐν γὰρ ἂν ἱκανὸς καὶ τέλεος καὶ πᾶσι φυτοῖς καὶ ζῴοις αἱρετὸς, οἷσπερ δυνατὸν ἐν οὕτως ἀεὶ διὰ βίου ζῆν]. And if one of us chose some other life, he would do so involuntarily, against what is truly choiceworthy by nature, out of ignorance or some unfortunate necessity.

Protarchus: That seems right.

Socrates: Then enough has been said, it seems to me, to show that the god of Philebus and the good cannot be recognized as one and the same.

Philebus: Nor is your intelligence the good, Socrates [οὐδὲ γὰρ σὸς νοῦς, ὦ Σώκρατες, ἔστι τἀγαθόν], since the same complaint applies to it.

Socrates: It may apply to *my* intelligence, Philebus, but not, I think, to the true and divine intelligence, which is in a quite different condition [οὐ μέντοι τόν γε ἀληθινὸν ἅμα καὶ θεῖον οἶμαι νοῦν, ἀλλ' ἄλλως πως ἔχειν]. (22a7-c6)

Socrates makes several important claims in this passage, and at first blush they weigh heavily in favor of the liberal interpretation. First, Socrates suggests that the CA's conclusion is grounded in the claim, accepted by Protarchus, that neither of the two unmixed lives is choiceworthy for any animal capable of living them. Thus Socrates seems to assume that Protarchus, at the crucial stages of the argument, justifiably rejects both lives from the standpoint of animal nature, and not just from the standpoint of human nature. This tells against my interpretation of the argument, which has it that—for Socrates, anyway—Protarchus is justified in rejecting the two lives only because neither of them is sufficiently human. Moreover, and again contrary to my interpretation, Socrates seems to assume that Protarchus is indeed capable of living both of the two rejected lives. If Socrates did not think this, then it is hard to see why he takes the CA to have shown that neither of the two unmixed lives is choiceworthy "for all the plants and animals that are able to live in this way consistently throughout their lives." Taken together, these two considerations strongly favor the liberal interpretation. For Socrates

seems to be saying that the two unmixed lives are to be rejected by Protarchus because each is missing something of final value, not because neither is livable by Protarchus *qua* human.

But things are not as clear as they seem. When Philebus uncharacteristically inserts himself into the conversation and reminds Socrates that the CA rules against pleasure and intelligence equally, Socrates demurs. He asserts that there is an important distinction between "his" intelligence and the "divine" intelligence, and implies that this distinction somehow redeems his earlier claim that intelligence is the good. He does not elaborate on the point, however, and we are left guessing at what he might mean. It is plausible to suppose that he takes the life of divine intelligence to be unlike the life of human intelligence in that it is "sufficient, complete, and choiceworthy" for all beings capable of living it. After all, this is precisely where the contrast would seem to matter. Now I take it that there are two different things he might mean by this. He might mean (A) that the unmixed life of *divine* intelligence, unlike the unmixed life of *human* intelligence, is sufficient, complete, and choiceworthy for all animals; or (B) that the unmixed life of divine intelligence is *not* sufficient, complete, and choiceworthy for all animals, *but only because mortals are not capable of living it*. If (A) is the right reading, then Socrates is claiming that, should the unmixed life of divine intelligence be put to the same choice test, it would not fail. But if (B) is the right reading, then he is claiming that the unmixed life of divine intelligence *would* fail the choice test—though only because the mortals to whom the choice is presented cannot, *qua* mortals, live it.

Perhaps the best way to settle the issue between (A) and (B) is to discover what Socrates thinks the divine life is like. Though he says relatively little about gods in the *Philebus*, there is a somewhat lengthy passage in which he suggests that their lives are both pleasureless and painless. This passage covers part of Socrates' attempt to spell out some of the consequences of the equilibrium model, which both he and Protarchus have by now accepted. According to Socrates, it is important to keep in mind that there is a kind of life that involves neither the restoration nor the destruction of equilibrium states. If the equilibrium model is correct, he thinks, then whoever lives such a life undergoes neither pleasure nor pain:

> Socrates: It was agreed before, in the comparison of lives, that the one who chooses the life of intelligence and reason must not experience pleasure, whether large or small.
> Protarchus: That was certainly agreed.
> Socrates: So [the one who chooses this life] lives in this way. And perhaps it would not be strange if, of all lives, this is the most divine [ἴσως οὐδὲν ἄτοπον εἰ πάντων τῶν βίων ἐστὶ θειότατος].
> Protarchus: Well, it's not likely that gods experience either pleasure or pain.
> Socrates: It certainly isn't. For each of these is inappropriate [ἄσχημον]. (33a8-b11)

By "inappropriate" here Socrates presumably means "inappropriate *for gods*." After all, the equilibrium model suggests that pleasure and pain are symptoms of an imperfect nature.[132] It is a standard feature of Platonic theology not only that gods have no such imperfections, but also that they are maximally intelligent. So it makes good sense

[132] Carone (2000, 262-264) accepts this point, but also holds that not all pleasures can be analyzed under EM. On her interpretation, the gods can have *pure* pleasures because these do not presuppose any defect in

for Socrates to point out here that the divine life features reason to the highest degree, but is free of both pleasure and pain. What makes this claim interesting for our purposes is that Socrates couches it in an implicit reference to the choice test. Indeed, he claims that the "most divine" life he is discussing here is none other than the pleasureless life of intelligence that he presented to Protarchus in the CA. And this suggests that, for Socrates anyway, it is not necessary to add the word "divine" to "the unmixed life of intelligence" when characterizing the life of the gods. There is nothing special in this context about the *intelligence* in a god's life; what makes a god's life special here is its freedom from the imperfections of embodiment, and the pleasure and pain that come along with them.

If this is the right way to understand the passage, then it seems clear that (A) is untenable. According to (A), remember, the unmixed life of *divine* intelligence, unlike the unmixed life of *human* intelligence, is sufficient, complete, and choiceworthy for all animals. But as we have seen, there is no meaningful distinction between these two lives; the divine life *just is* the unmixed life of intelligence. And Socrates cannot mean that such a life is sufficient, complete, and choiceworthy for all animals, since the second part of the CA purports to establish that this is not the case. This leaves us with (B), according to which the unmixed life of intelligence is not choiceworthy for all animals, since mortals are incapable of living such a life.[133] The main force of (B), remember, is that it turns the chooser's ability to live a given life into a necessary condition for that life's being

---

the pleasure-taker. As I have already argued, however, the pure pleasures are best understood as restorations of equilibrium states. See above, section 1.13.

choiceworthy for that chooser. Indeed, this is presumably the main reason why Socrates is willing to concede that the unmixed life of intelligence—that is, the divine life—is not choiceworthy for us.

With this new necessary condition in mind, let us return for a moment to what Socrates originally says about the force of the CA. In his view, remember, the argument shows that neither of the two unmixed lives is the good, since if either of them were, "it would be sufficient and complete and choiceworthy for all the plants and animals that are able to live in this way consistently throughout their lives." Now according to the liberal interpretation, Socrates is implying here that Protarchus—the animal to whom the choice has been presented in the CA—is able to live both of the unmixed lives, since otherwise the CA would not have shown what Socrates takes it to have shown. But according to the best interpretation of what Socrates says about "divine intelligence" shortly thereafter, Socrates thinks that Protarchus, *qua* human, is not able to live the unmixed life of intelligence. It is important to notice, however, that this interpretation does not invalidate the results of the CA. Instead, it requires adding a necessary condition for a life's being choiceworthy for any given agent. And this necessary condition can be captured by altering the liberal interpretation of Socrates' remarks about the force of the CA.

According to the liberals, remember, Socrates thinks that the choice test of a given life is inconclusive if the choosing agent is not capable of living it. But on my reading, the choice test of a given life yields a *negative* conclusion if the choosing agent is not capable of living it. Put simply, this is a dispute about how to interpret the logical form of the

[133] Cf. Frede (1992, 453).

choice test. The liberal interpretation has it that Socrates guarantees the chooser's ability to live the life by restricting the domain of the quantifier:

> For any animal A capable of living some life L,
>
> > if L is the good life for all animals, then L is sufficient, complete, and choiceworthy for A.

On this interpretation, Socrates is implying that if the choosing agent cannot live the life with which he is being presented, then the choice test cannot arrive at a negative result about that life. But there is another way to go. We can add a quantifier ranging over lives—along with a specific necessary condition on the choiceworthiness of a life—such that the choice test yields a negative result if the chooser is unable to live the life in question:

> For any animal A and any life L,
>
> > if L is the good life for all animals, then L is sufficient, complete, and choiceworthy for A. And
>
> > if L is choiceworthy for A, then A is capable of living L.

This interpretation, unlike the former, allows Socrates to generate a negative conclusion if the life being presented to the chooser is one that the chooser cannot live. It also makes excellent sense of Socrates' "divine intelligence" caveat. What he wants to say, I take it, is that the negative result of the CA need not apply equally to both gods and mortals. It may be true that the unmixed life of intelligence is not choiceworthy for

mortals, but from this it does not follow that a life of unmixed intelligence is not choiceworthy for gods. So when L is the life of unmixed intelligence, the test (as I interpret it) yields a negative result when presented to mortals, but no result when presented to gods.[134] The liberal interpretation, on the other hand, cannot produce these results. For its version of the choice test yields no conclusion for the unmixed life of intelligence when that life is presented to a mortal chooser. As we have already seen, however, this is contrary to the result that Socrates takes himself to have established in the second part of the CA. So in fact it is the liberal interpretation of the passage which cannot make sense of the CA as Socrates himself understands it.

The upshot of all this is that the CA is designed to establish much less than is often assumed. The liberals suppose that the point of the argument is to show that some types of intelligence and some types of pleasure are final goods. But as we have seen, their interpretation of the choice test does not fit the flow of the argument. To see the shape of my alternative more clearly, consider the following more robust and textually grounded interpretation of the choice test—as modified by, among other things, the added necessary condition that a life is choiceworthy for an agent only if that agent can live it:

> For any animal A, and for any life L,
>
> if L is the good life for all animals, then L is sufficient, complete, and choiceworthy for A (22b3-6).

[134] If Socrates' position is to cohere, the choice test should yield only negative results, and only those negative results that suit the needs of the argument. And what the argument needs is to turn away both the unmixed life of pleasure and the unmixed life of intelligence when humans are presented with the choice. Everything else should remain open.

(1) If L is sufficient and complete, then L contains all the goods there are (20e4-21a2).

(2) If L is choiceworthy for A, then
if A chooses rationally, knowledgeably, and voluntarily[135] (22b6-8), then
(i) A is capable of living L (22b5-6, 22c5-6) and
(ii) A chooses L. (22b6-8)

If this is a fair rendering of the choice test, then there are three different ways in which L might fail it. First, (1) might be false because L is missing some goods. Second, (2i) might be false because A cannot live L. And third, (2ii) might be false because A fails to choose L. What this means, I take it, is that one cannot infer from a life's failure to pass the choice test that (1) is false. This life might have all the final goods there are but still not be livable by the agent. As I suggested earlier, then, the CA does not license any robust conclusion about whether there are any genuinely good types of pleasure or cognition. This is not its point. The CA is designed to reach a twofold negative conclusion about the *good life for mortals*—namely, that it is neither wholly unmixed with pleasure nor wholly unmixed with intelligence. The CA is *not* designed to explain *why* neither of these lives is the good one. In my view, Socrates deliberately leaves this question open until later, when he advances arguments (including the PA) to the effect that no pleasures are good in themselves.

[135] As I have articulated it, the livability condition on choiceworthiness has bite only if the chooser is rational. So if the chooser is a mollusk, then the mollusk's inability to lead a philosopher's life does not yield the conclusion that the philosopher's life is not the good one for animals—given, of course, that mollusks are fundamentally non-rational. On the other hand, the philosopher's inability to lead a mollusk's life *does* yield the desired negative conclusion about the *mollusk's* life.

3.11. Conclusion

If my interpretation is correct, then there are no adequate exegetical grounds to dispute the claim that Plato is a radical anti-hedonist. The conclusion of the CA is entirely consistent with the conclusion of the PA, and in the latter argument Plato simply denies that any pleasure can be worth pursuing for its own sake. As we have seen, this radical position stems from the thought, articulated most clearly in the PT, that pleasures are things that *essentially aim*. And in Plato's view, to place final value on a thing that essentially aims is to lose sight of that thing's target, and hence to lose sight of both its nature and its value. On the other hand, Plato also denies that the unmixed life of intelligence is best for us; in his view, the best life for us is a rather pleasant one. But what accounts for the superlative value of this life is not its pleasantness, according to Plato, since no pleasures can be bearers of final value. They are included in the best human life for some other reason, either because they are complements of cognitive activity, or because they are vital modes of perceiving final value.[136]

The crux of Plato's position, then, is that acknowledging pleasure's rational, deliberative authority comes at a price for the hedonist. For if its function is to *pick out* objects worthy of ultimate choice, then it itself cannot *be* an object of ultimate choice. This consideration brings Plato to an account of pleasure's value that is both radically

anti-hedonist and firmly anti-ascetic. In his view, pleasure is a fallible yet indispensable *guide*, and for this reason alone it cannot qualify as a *goal*.

[136] Though I suspect that both of these considerations are in play, I do not wish to insist on that here. I want only to emphasize that Plato includes pleasures in the good life for some reason other than their being final goods.

Bibliography

Annas, J. (1999) *Platonic Ethics, Old and New*. Ithaca: Cornell University Press.

Anscombe, G. E. M. (1957) *Intention*. Ithaca: Cornell University Press.

Benitez, E. E. (1989) *Forms in Plato's* Philebus. Assen: Van Gorcum.

Bobonich, C. (1995) "Plato's Theory of Goods in the *Laws* and *Philebus*" in *Proceedings of the Boston Area Colloquium in Ancient Philosophy* 11: 101-136.

Bobonich, C. (2002) *Plato's Utopia Recast: His Later Ethics and Politics*. Oxford: Clarendon Press.

Bury, R. (1897) *The* Philebus *of Plato*. Cambridge: Cambridge University Press.

Carone, G. (2000) "Hedonism and the Pleasureless Life in Plato's *Philebus*," *Phronesis* 45: 257-283.

Caston, V. (1993) "Towards a History of the Problem of Intentionality Among the Greeks," *Proceedings of the Boston Area Colloquium in Ancient Philosophy* 9: 213-245.

Cooper, J. (1999) *Reason and Emotion*. Princeton: Princeton University Press.

Davidson, D. (1980) *Essays on Actions and Events*. Oxford: Clarendon Press.

Davidson, D. (1990) *Plato's Philebus*. New York: Garland Publishing.

Delcomminette, S. (2003) "False Pleasures, Appearance and Imagination in the *Philebus*," *Phronesis* 48: 215-237.

Dybikowski, J. C. (1970a) "False Pleasure and the *Philebus*," *Phronesis* 15: 147-165.

Dybikowski, J. C. (1970b) "Mixed and False Pleasure in the *Philebus*: A Reply," *Philosophical Quarterly* 20: 244-247.

Frede, D. (1985) "Rumpelstilskin's Pleasures: True and False Pleasures in Plato's *Philebus*," *Phronesis* 30: 151-180.

Frede, D. (1992) "Disintegration and Restoration: Pleasure and Pain in Plato's *Philebus*," in R. Kraut, ed., *The Cambridge Companion to Plato*. Cambridge: Cambridge University Press.

Frede, D. (1993) *Plato:* Philebus*. Translated, with Introduction & Notes*. Indianapolis: Hackett.

Gallop, D. (1960) "True and False Pleasures," *Philosophical Quarterly* 10: 331-342.

Gocer, A. (1999) "Hesuchia, a Metaphysical Principle in Plato's Moral Psychology," *Apeiron* 32: 17-36.

Gosling, G. C. B. (1959) "False Pleasures: *Philebus* 35c-41b," *Phronesis* 4: 44-53.

Gosling, G. C. B.(1975) *Plato:* Philebus*. Translated with Notes and Commentary*. Oxford: Clarendon Press.

Gosling, G. C. B. and C. C. W. Taylor (1982) *The Greeks on Pleasure*. Oxford: Clarendon Press.

Griswold, C. L. (1988) Ed. *Platonic Writings, Platonic Readings*. New York: Routledge.

Hackforth, R. (1972) *Plato's* Philebus*: Translated with an Introduction and Commentary*. Cambridge: Cambridge University Press.

Hampton, C. (1990) *Pleasure, Knowledge, and Being: An Analysis of Plato's* Philebus*.* Albany: SUNY Press.

Irwin, T. (1995) *Plato's Ethics*. Oxford: Oxford University Press.

Kenny, A. (1960) "False Pleasures in the *Philebus*: A Reply to Mr. Gosling," *Phronesis* 5: 45-52.

Lemos, N. (1994) *Intrinsic Value: Concept and Warrant*. Cambridge: Cambridge University Press.

Lovibond, S. (1990) "True and False Pleasures," *Proceedings of the Aristotelian Society* 90: 213-230.

Millgram, E. (1993) "Pleasure in Practical Reasoning," *The Monist* 90: 394-415.

Mills, M. J. (1985) "ΦΘΟΝΟΣ and its related ΠΑΘΗ in Plato and Aristotle," *Phronesis* 30: 1-12.

Moordian, N. (1996) "Converting Protarchus: Relativism and False Pleasures of Anticipation in Plato's *Philebus*," *Ancient Philosophy* 16: 93-112.

Moore, G. E. (1993) *Principia Ethica: Revised Edition*. T. Baldwin, Ed. Cambridge: Cambridge University Press.

Nagel, T. (1986) *The View From Nowhere*. Oxford: Oxford University Press.

Penelhum, T. (1964) "Pleasure and Falsity," *American Philosophical Quarterly* 1: 81-91.

Penner, T. (1970) "False Anticipatory Pleasures: *Philebus* 36a3-41a6," *Phronesis* 15: 166-178.

Sayre, K. (1983) *Plato's Late Ontology: A Riddle Resolved*. Princeton: Princeton University Press.

Shiner, R. (1974) *Knowledge and Reality in Plato's* Philebus. Assen: Van Gorcum.

Thalberg, I. (1962) "False Pleasures," *Journal of Philosophy* 59: 65-74.

Tuozzo, T. (1996) "The General Account of Pleasure in Plato's *Philebus*," *Journal of the History of Philosophy* 34: 495-513.

Van Riel, G. (2000) *Pleasure and the Good Life*. Leiden: Brill.

Westerlink, L. G. (1959) *Damascius: Lectures on the* Philebus*. Text, Translation, Notes, and Indices*. Amsterdam: North-Holland Publishing Company.

White, N. (2002) *Individual and Conflict in Greek Ethics*. Oxford: Clarendon Press.

Williams, B. (1959) "Pleasure and Belief," *Proceedings of the Aristotelian Society* Supplementary Volume 33: 57-72.

Vita

Matthew Lyall Evans was born in Washington, D.C. on May 20, 1970, the son of Mark and Janet Evans. After finishing high school at St. Albans in 1988, he entered Vassar College, from which he eventually received the degree of Bachelor of Arts. After several years of intensive wandering, he joined the University of Texas Philosophy Department as a graduate student in 1996. In 1998 he received a Master of Arts degree from the University of Texas.

Permanent address: 1409 Park Row, La Jolla, CA 92037

This dissertation was typed by Matthew Lyall Evans.